

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
18 July 2002 (18.07.2002)

PCT

(10) International Publication Number  
**WO 02/055658 A2**

(51) International Patent Classification: C12N

(21) International Application Number: PCT/US01/50638

(22) International Filing Date: 29 October 2001 (29.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/244,685 30 October 2000 (30.10.2000) US

(71) Applicant: EXELIXIS PLANT SCIENCES, INC.  
[US/US]; 16160 S.W. Upper Boones Ferry Road, Portland,  
OR 97224-7744 (US).

(72) Inventors: CONNERS, Karin; 16516 S.W. Jesse Court,  
Aloha, OR 97007 (US). MATHEWS, Helena, V.; 14546  
N.W. Joseph Court, Portland, OR 97229 (US). LIU, Alex;  
2061 W. 16th Avenue, Eugene, OR 97402 (US).

(74) Agents: BRUNELLE, Jan et al.; Exelixis, Inc., 170 Har-  
bor Way, South San Francisco, CA 94083-0511 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EE, ES, FI, GE, GH, GM, HR, HU,  
ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS,  
LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO,  
NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR,  
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

**Published:**

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: IDENTIFICATION AND CHARACTERIZATION OF AN ANTHOCYANIN MUTANT (ANT1) IN TOMATO

(57) Abstract: The present invention is directed to a novel plant phenotype, designated *Anthocyanin 1* (ANT1), a nucleic acid se-  
quence expressed in plants demonstrating the *ANT1* phenotype and the corresponding amino acid sequence. Also provided are plant  
cells that exhibit modified *ANT1* expression.

WO 02/055658 A2

## IDENTIFICATION AND CHARACTERIZATION OF AN ANTHOCYANIN MUTANT (*ANT1*) IN TOMATO

### REFERENCE TO RELATED APPLICATIONS

[0001] This Application claims priority to U.S. provisional patent application serial No. 60/244,685, filed October 30, 2000, the contents of which are hereby incorporated in its entirety.

### FIELD OF THE INVENTION

[0002] The present invention relates to a plant phenotype, designated *Anthocyanin 1* (*ANT1*), together with DNA and polypeptide sequences associated with the same.

### BACKGROUND OF THE INVENTION

[0003] The traditional methods for gene discovery, including chemical mutagenesis, irradiation and T-DNA insertion, used to screen loss of function mutants have limitations. Mutagenic methods such as these rarely identify genes that are redundant in the genome, and gene characterization is time-consuming and laborious.

[0004] Activation tagging is a method by which genes are randomly and strongly up-regulated on a genome-wide scale, after which specific phenotypes are screened for and selected. Isolation of mutants by activation tagging has been reported (Hayashi *et al.*, 1992). An activation T-DNA tagging construct was used to activate genes in tobacco cell culture allowing the cells to grow in the absence of plant growth hormones (Walden *et al.*, 1994). Genes have been isolated from plant genomic sequences flanking the T-DNA tag and putatively assigned to plant growth hormone responses. (See, *e.g.*, Miklashevichs *et al.* 1997, Harling *et al.*, 1997; Walden *et al.*, 1994; and Schell *et al.*, 1998, which discusses related studies.)

[0005] The first gene characterized in *Arabidopsis* using activation tagging was a gene encoding the histone kinase involved in the cytokinin signal transduction pathway. The gene sequence was isolated from plant genomic DNA by plasmid rescue and the role of the gene, *CK11*, in cytokinin responses in plants was confirmed by re-introduction into *Arabidopsis* (Kakimoto, 1996). This was followed by reports of several dominant mutants such as TINY, LHY and SHI using a similar approach along with the Ds transposable element (Wilson *et al.*, 1996, Schaffer *et al.*, 1998, Fridborg

*et al.*, 1999). In a more recent report, activation T-DNA tagging and screening plants for an early flowering phenotype led to the isolation of the FT gene (Kardailsky *et al.*, 1999).

[0006] The potential application of activation tagging as a high through put technology for gene discovery has been demonstrated based on screening of several dominant mutant genes involved in photoreceptor, brassinosteroid, gibberellin and flowering signal pathways, as well as disease resistance. (See, *e.g.*, Weigel *et al.*, 2000, Christensen *et al.*, 1998; Kardailsky *et al.*, 1999).

[0007] *Arabidopsis* has been widely used as a model for plant improvement for plants such as *Brassica* species having a siliques type of fruit. However, *Arabidopsis* does not serve as a model for plants having a fleshy fruit.

[0008] A method for identifying and characterizing genes based on modified gene expression in fruit-bearing plants is described in PCT publication WO0053794. Dwarf varieties of fruit-bearing plants, particularly dwarf varieties of tomato, are useful in the overexpression of one or more native plant genes and in correlating that overexpression with a particular phenotype.

[0009] Dwarf tomatoes are characterized by their short internodes, which give plants a compact appearance. The miniature *Lycopersicon esculentum* cultivar, Micro-Tom is a proportionally dwarfed plant that grows at high density (up to 1357 plants/m<sup>2</sup>), has a short life cycle (70-80 days from sowing to fruit ripening), and for which fruit size, and leaf size have been genetically reduced (Meissner *et al.*, 1997; Scott and Harbaugh, 1989). In addition, Micro-Tom has been shown to be resistant to a number of diseases and can be transformed at frequencies of up to 80% through *Agrobacterium*-mediated transformation of cotyledons (Meissner *et al.*, 1997). Similar to Micro-Tom, Florida Petite (Fla. Agr. Expt. Sta. Circ. S-285), Tiny Tim and Small Fry are dwarf varieties of tomato which have a short life cycle, and for which fruit size, and leaf size have been genetically reduced.

[0010] Efforts are underway in industry and academia to develop a means to identify genes associated with particular plant traits or characteristics in order to develop improved plants having such traits. The present invention provides a plant phenotype associated with modified expression of a native plant gene.

[0011] In an activation tagging screen in Micro-Tom, we identified a gene involved in pigment production. Anthocyanins are pigments that are responsible for many of the red and blue colors in plants. The genetic basis of anthocyanin biosynthesis has been

well characterized in corn, Petunia, and Antirrhinium (Dooner et al, 1991; Jayaram and Peterson, 1990; Quattrocchio F et al., 1999).

### **SUMMARY OF THE INVENTION**

- [0012] The invention provides nucleic acid and amino acid sequences associated with the *Anthocyanin 1* ("*ANT1*") phenotype in plants, presented as modified leaf, flower or fruit color.
- [0013] In one aspect, the invention provides one or more isolated *ANT1* nucleic acid sequences comprising a nucleic acid sequence that encodes or is complementary to a sequence that encodes an *ANT1* polypeptide having at least 70%, 80%, 90% or more sequence identity to the amino acid sequence presented as SEQ ID NO:2.
- [0014] In another aspect, the polynucleotide comprises a nucleic acid sequence that hybridizes, under high, medium, or low stringency conditions to the nucleic acid sequence, or fragment thereof, presented as SEQ ID NO:1, or the complement thereof.
- [0015] In a related aspect, expression of one or more of such *ANT1* polynucleotides in a plant is associated with the *ANT1* phenotype.
- [0016] The invention further provides plant transformation vectors, plant cells, plant parts and plants comprising an *ANT1* nucleic acid sequence.
- [0017] Expression of such an *ANT1* nucleic acid sequence in a plant is associated with the *ANT1* phenotype, presented as a modified leaf, flower or fruit color phenotype.
- [0018] The expression of an *ANT1* nucleic acid sequence may be modified in ornamental plants, fruit and vegetable-producing plants, grain-producing plants, oil-producing plants and nut-producing plants, as well as other crop plants, resulting in the *ANT1* phenotype.
- [0019] In a further aspect the invention provides a method of modifying the *ANT1* phenotype in a plant by introducing an *ANT1* nucleic acid sequence into plant progenitor cells and growing the cells to produce a transgenic plant.

### **DETAILED DESCRIPTION OF THE INVENTION**

#### **Definitions**

- [0020] Unless otherwise indicated, all technical and scientific terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook *et al.*, 1989, and Ausubel FM *et al.*, 1993, for definitions and terms of the art. It is to be understood that this invention is

not limited to the particular methodology, protocols, and reagents described, as these may vary.

[0021] All publications cited herein, and listed below immediately after the examples, are expressly incorporated herein by reference for the purpose of describing and disclosing compositions and methodologies that might be used in connection with the invention. All cited patents, patent publications, and sequence and other information in referenced websites are also incorporated by reference.

[0022] As used herein, the term "vector" refers to a nucleic acid construct designed for transfer between different host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

[0023] A "heterologous" nucleic acid construct or sequence has a portion of the sequence which is not native to the plant cell in which it is expressed. Heterologous, with respect to a control sequence refers to a control sequence (*i.e.* promoter or enhancer) that does not function in nature to regulate the same gene the expression of which it is currently regulating. Generally, heterologous nucleic acid sequences are not endogenous to the cell or part of the genome in which they are present, and have been added to the cell, by infection, transfection, microinjection, electroporation, or the like. A "heterologous" nucleic acid construct may contain a control sequence/DNA coding sequence combination that is the same as, or different from a control sequence/DNA coding sequence combination found in the native plant.

[0024] As used herein, the term "gene" means the segment of DNA involved in producing a polypeptide chain, which may or may not include regions preceding and following the coding region, *e.g.* 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons).

[0025] As used herein, "percent (%) sequence identity" with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of nucleotides or amino acids in the candidate derivative sequence identical with the nucleotides or amino acids in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J.

Mol. Biol. (1997) 215:403-410; blast.wustl.edu/blast/README.html website) with all the search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A % identity value is determined by the number of matching identical nucleotides or amino acids divided by the sequence length for which the percent identity is being reported. "Percent (%) amino acid sequence similarity" is determined by doing the same calculation as for determining % amino acid sequence identity, but including conservative amino acid substitutions in addition to identical amino acids in the computation.

[0026] The term "% homology" is used interchangeably herein with the term "% identity."

[0027] A nucleic acid sequence is considered to be "selectively hybridizable" to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Hybridization conditions are based on the melting temperature ( $T_m$ ) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about  $T_m - 5^\circ\text{C}$  ( $5^\circ$  below the  $T_m$  of the probe); "high stringency" at about  $5 - 10^\circ$  below the  $T_m$ ; "intermediate stringency" at about  $10 - 20^\circ$  below the  $T_m$  of the probe; and "low stringency" at about  $20 - 25^\circ$  below the  $T_m$ . Functionally, maximum stringency conditions may be used to identify sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify sequences having about 80% or more sequence identity with the probe.

[0028] Moderate and high stringency hybridization conditions are well known in the art (see, for example, Sambrook, *et al.*, 1989, Chapters 9 and 11, and in Ausubel, F.M., *et al.*, 1993, expressly incorporated by reference herein). An example of high stringency conditions includes hybridization at about  $42^\circ\text{C}$  in 50% formamide, 5X SSC, 5X Denhardt's solution, 0.5% SDS and 100  $\mu\text{g/ml}$  denatured carrier DNA followed by washing two times in 2X SSC and 0.5% SDS at room temperature and two additional times in 0.1X SSC and 0.5% SDS at  $42^\circ\text{C}$ .

[0029] As used herein, "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid sequence or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the

cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention.

**[0030]** As used herein, the terms "transformed", "stably transformed" or "transgenic" with reference to a plant cell means the plant cell has a non-native (heterologous) nucleic acid sequence integrated into its genome which is maintained through two or more generations.

**[0031]** As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation.

**[0032]** The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell where the nucleic acid sequence may be incorporated into the genome of the cell (for example, chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA).

**[0033]** As used herein, a "plant cell" refers to any cell derived from a plant, including cells from undifferentiated tissue (*e.g.*, callus) as well as plant seeds, pollen, propagules and embryos.

**[0034]** As used herein, the terms "native" and "wild-type" relative to a given plant trait or phenotype refers to the form in which that trait or phenotype is found in the same variety of plant in nature.

**[0035]** As used herein, the term "modified" regarding a plant trait, refers to a change in the phenotype of a transgenic plant relative to a non-transgenic plant, as it is found in nature.

**[0036]** As used herein, the term " $T_1$ " refers to the generation of plants from the seed of  $T_0$  plants. The  $T_1$  generation is the first set of transformed plants that can be selected by application of a selection agent, *e.g.*, an antibiotic or herbicide, for which the transgenic plant contains the corresponding resistance gene.

**[0037]** As used herein, the term " $T_2$ " refers to the generation of plants by self-fertilization of the flowers of  $T_1$  plants, previously selected as being transgenic.

**[0038]** As used herein, the term "plant part" includes any plant organ or tissue including, without limitation, seeds, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. Plant cells can be

obtained from any plant organ or tissue and cultures prepared therefrom. The class of plants which can be used in the methods of the present invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledenous and dicotyledenous plants.

[0039] As used herein, "transgenic plant" includes reference to a plant that comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic.

[0040] Thus a plant having within its cells a heterologous polynucleotide is referred to herein as a "transgenic plant". The heterologous polynucleotide can be either stably integrated into the genome, or can be extra-chromosomal. Preferably, the polynucleotide of the present invention is stably integrated into the genome such that the polynucleotide is passed on to successive generations. The polynucleotide is integrated into the genome alone or as part of a recombinant expression cassette. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acids including those transgenics initially so altered as well as those created by sexual crosses or asexual reproduction of the initial transgenics.

[0041] A plant cell, tissue, organ, or plant into which the recombinant DNA constructs containing the expression constructs have been introduced is considered "transformed", "transfected", or "transgenic". A transgenic or transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of a recombinant nucleic acid sequence. Hence, a plant of the invention will include any plant which has a cell containing a construct with introduced nucleic acid sequences, regardless of whether the sequence was introduced into the directly through transformation means or introduced by generational transfer from a progenitor cell which originally received the construct by direct transformation.



- [0042] The terms "*Anthocyanin 1*" and "*ANT1*", as used herein encompass native *Anthocyanin 1* (*ANT1*) nucleic acid and amino acid sequences, homologues, variants and fragments thereof.
- [0043] An "isolated" *ANT1* nucleic acid molecule is an *ANT1* nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the *ANT1* nucleic acid. An isolated *ANT1* nucleic acid molecule is other than in the form or setting in which it is found in nature. However, an isolated *ANT1* nucleic acid molecule includes *ANT1* nucleic acid molecules contained in cells that ordinarily express *ANT1* where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.
- [0044] As used herein, the term "mutant" with reference to a polynucleotide sequence or gene differs from the corresponding wild type polynucleotide sequence or gene either in terms of sequence or expression, where the difference contributes to a modified plant phenotype or trait. Relative to a plant or plant line, the term "mutant" refers to a plant or plant line which has a modified plant phenotype or trait, where the modified phenotype or trait is associated with the modified expression of a wild type polynucleotide sequence or gene.
- [0045] Generally, a "variant" polynucleotide sequence encodes a "variant" amino acid sequence which is altered by one or more amino acids from the reference polypeptide sequence. The variant polynucleotide sequence may encode a variant amino acid sequence having "conservative" or "non-conservative" substitutions. Variant polynucleotides may also encode variant amino acid sequences having amino acid insertions or deletions, or both.
- [0046] As used herein, the term "phenotype" may be used interchangeably with the term "trait". The terms refer to a plant characteristic which is readily observable or measurable and results from the interaction of the genetic make-up of the plant with the environment in which it develops. Such a phenotype includes chemical changes in the plant make-up resulting from enhanced gene expression which may or may not result in morphological changes in the plant, but which are measurable using analytical techniques known to those of skill in the art.
- [0047] As used herein, the term "interesting phenotype" with reference to a plant produced by the methods described herein refers to a readily observable or measurable phenotype demonstrated by a T<sub>1</sub> and/or subsequent generation plant, which is not

displayed by a plant that has not been so transformed (and/or is not the progeny of a plant that has been so transformed) and represents an improvement in the plant. An "improvement" is a feature that may enhance the utility of a plant species or variety by providing the plant with a unique quality. By unique quality is meant a novel feature or a change to an existing feature of the plant species which is a quantitative change (increase or decrease) or a qualitative change in a given feature or trait.

#### **The Identified *ANTI* Phenotype and Gene**

- [0048] The gene and phenotype of this invention were identified in a screen using activation tagging. Activation tagging is a process by which a heterologous nucleic acid construct comprising a nucleic acid control sequence, *e.g.* an enhancer, is inserted into a plant genome. The enhancer sequences act to enhance transcription of a one or more native plant genes (See, *e.g.*, Walden R, *et al.*, 1994; Weigel D *et al.* 2000).
- [0049] Briefly, a large number of tomato (*Lycopersium esculentum*) cv. Micro-Tom plants were transformed with a modified form of the activation tagging vector pSKI015 (Weigel *et al.*, 2000), which comprises a T-DNA (*i.e.*, the sequence derived from the Ti plasmid of *Agrobacterium tumefaciens* that are transferred to a plant cell host during *Agrobacterium* infection), an enhancer element and a selectable marker gene. The construct, pAG3202, is further described in the Examples. Following random insertion of pAG3202 into the genome of transformed plants, the enhancer element can result in up-regulation genes in the vicinity of the T-DNA insertion, generally within 5-10 kilobase (kb) of the insertion. In the T<sub>1</sub> generation, plants were exposed to the selective agent in order to specifically recover those plants that expressed the selectable marker and therefore harbored insertions of the activation-tagging vector. Transformed plants were observed for interesting phenotypes, which are generally identified at the T<sub>1</sub>, T<sub>2</sub> and/or T<sub>3</sub> generations. Interesting phenotypes may be identified based on morphology, a biochemical screen, herbicide tolerance testing, herbicide target identification, fungal or bacterial resistance testing, insect or nematode resistance testing, screening for stress tolerance, such as drought, salt or antibiotic tolerance, and output traits, such as oil, starch, pigment, or vitamin composition. Genomic sequence surrounding the T-DNA insertion is analyzed in order to identify genes responsible for the interesting phenotypes. Genes responsible for causing such phenotypes are identified as attractive targets for manipulation for agriculture, food, ornamental plant, and/or pharmaceutical industries.

- [0050] It will be appreciated that in most cases when a modified phenotype results from the enhanced expression of a tagged gene, the phenotype is dominant. In some cases, the enhanced expression of a given native plant gene or a fragment thereof may result in decreased expression or inactivation of its homologue or another native plant gene, which results in the interesting phenotype. The T-DNA insertion may also result in disruption ("loss-of-function") of a native plant gene, in which case the phenotype is generally recessive.
- [0051] The present invention provides a modified leaf, flower or fruit color phenotype, identified in ACTTAG Mico-Tom lines that were observed at the callus stage as having purple color and purple shoots. Purple plants were derived from purple colored caulogenic callus in culture. The clonal plant lines (i.e., additional shoots originating from the same purple colored caulogenic callus or those multiplied from the first purple plant either in tissue culture or by cuttings in the greenhouse) were identified as having purple coloration on leaves, sepals and flowers. The plants were also observed to exhibit a modified fruit color described as a deeper red color relative to wild type Micro-Tom plants. The phenotype and associated gene have been designated *Anthocyanin 1* ("*ANT1*").
- [0052] The invention also provides a newly identified and isolated nucleic acid sequence that was identified by analysis of the genomic DNA sequence surrounding the T-DNA insertion correlating with the *ANT1* phenotype. In particular, applicants have identified and characterized the open reading frame of the *ANT1* gene, which is specifically overexpressed in plants having the *ANT1* phenotype, and which is provided in SEQ ID NO:1. A detailed description of the isolation and characterization of *ANT1* is set forth in the Examples.

#### Compositions of the Invention

##### *ANT1* Nucleic acids

- [0053] The *ANT1* gene may be used in the development of transgenic plants having a desired phenotype. This may be accomplished using the native *ANT1* sequence, a variant *ANT1* sequence or a homologue or fragment thereof.
- [0054] A *ANT1* nucleic acid sequence of this invention may be a DNA or RNA sequence, derived from genomic DNA, cDNA or mRNA. The nucleic acid sequence may be cloned, for example, by isolating genomic DNA from an appropriate source, and amplifying and cloning the sequence of interest using PCR. Alternatively, nucleic acid

sequence may be synthesized, either completely or in part, especially where it is desirable to provide plant-preferred sequences. Thus, all or a portion of the desired structural gene (that portion of the gene which encodes a polypeptide or protein) may be synthesized using codons preferred by a selected host.

[0055] The invention provides a polynucleotide comprising a nucleic acid sequence which encodes or is complementary to a sequence which encodes an *ANT1* polypeptide having the amino acid sequence presented in SEQ ID NO:2 and a polynucleotide sequence identical over its entire length to the *ANT1* nucleic acid sequence presented SEQ ID NO:1. The invention also provides the coding sequence for the mature *ANT1* polypeptide, a variant or fragment thereof, as well as the coding sequence for the mature polypeptide or a fragment thereof in a reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, pro-, or prepro- protein sequence.

[0056] A *ANT1* polynucleotide can also include non-coding sequences, including for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, untranslated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence that encodes additional amino acids. For example, a marker sequence can be included to facilitate the purification of the fused polypeptide. Polynucleotides of the present invention also include polynucleotides comprising a structural gene and the naturally associated sequences that control gene expression.

[0057] When an isolated polynucleotide of the invention comprises an *ANT1* nucleic acid sequence flanked by non- *ANT1* nucleic acid sequence, the total length of the combined polynucleotide is typically less than 25 kb, and usually less than 20kb, or 15 kb, and in some cases less than 10 kb, or 5 kb.

[0058] In addition to the *ANT1* nucleic acid and corresponding polypeptide sequences described herein, it is contemplated that *ANT1* variants can be prepared. *ANT1* variants can be prepared by introducing appropriate nucleotide changes into the *ANT1* nucleic acid sequence; by synthesis of the desired *ANT1* polypeptide or by altering the expression level of the *ANT1* gene in plants. Those skilled in the art will appreciate that amino acid changes may alter post-translational processing of the *ANT1* polypeptide, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.

- [0059] In one aspect, preferred *ANTI* coding sequences include a polynucleotide comprising a nucleic acid sequence which encodes or is complementary to a sequence which encodes an *ANTI* polypeptide having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95% or more sequence identity to the amino acid sequence presented in SEQ ID NO:2.
- [0060] In another aspect, preferred variants include an *ANTI* polynucleotide sequence that is at least 50% to 60% identical over its entire length to the *ANTI* nucleic acid sequence presented as SEQ ID NO:1, and nucleic acid sequences that are complementary to such an *ANTI* sequence. More preferable are *ANTI* polynucleotide sequences comprise a region having at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the *ANTI* sequence presented as SEQ ID NO:1.
- [0061] In a related aspect, preferred variants include polynucleotides that are be "selectively hybridizable" to the *ANTI* polynucleotide sequence presented as SEQ ID NO:1.
- [0062] Sequence variants also include nucleic acid molecules that encode the same polypeptide as encoded by the *ANTI* polynucleotide sequence described herein. Thus, where the coding frame of an identified nucleic acid molecules is known, for example by homology to known genes or by extension of the sequence, it is appreciated that as a result of the degeneracy of the genetic code, a number of coding sequences can be produced. For example, the triplet CGT encodes the amino acid arginine. Arginine is alternatively encoded by CGA, CGC, CCG, AGA, and AGG. Therefore it is appreciated that such substitutions in the coding region fall within the sequence variants that are covered by the present invention. Any and all of these sequence variants can be utilized in the same way as described herein for the identified *ANTI* parent sequence, SEQ ID NO:1.
- [0063] It is further appreciated that such sequence variants may or may not selectively hybridize to the parent sequence. This would be possible, for example, when the sequence variant includes a different codon for each of the amino acids encoded by the parent nucleotide. Such variants are, nonetheless, specifically contemplated and encompassed by the present invention. In accordance with the present invention, also encompassed are sequences that at least 70% identical to such degeneracy-derived sequence variants.
- [0064] Although *ANTI* nucleotide sequence variants are preferably capable of hybridizing to the nucleotide sequences recited herein under conditions of moderately high or high

stringency, there are, in some situations, advantages to using variants based on the degeneracy of the code, as described above. For example, codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic organism, in accordance with the optimum codon usage dictated by the particular host organism. Alternatively, it may be desirable to produce RNA having longer half lives than the mRNA produced by the recited sequences.

[0065] Variations in the native full-length *ANT1* nucleic acid sequence described herein, may be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations, as generally known in the art, oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed mutagenesis (Kunkel TA *et al.*, 1991); cassette mutagenesis (Crameri A *et al.*, 1995); restriction selection mutagenesis (Haught C *et al.*, 1994), or other known techniques can be performed on the cloned DNA to produce nucleic acid sequences encoding *ANT1* variants.

[0066] It is contemplated that the gene sequences associated with the *ANT1* phenotype may be synthesized, either completely or in part, especially where it is desirable to provide host-preferred sequences. Thus, all or a portion of the desired structural gene (that portion of the gene which encodes the protein) may be synthesized using codons preferred by a selected host. Host-preferred codons may be determined, for example, from the codons used most frequently in the proteins expressed in a desired host species.

[0067] It is preferred that an *ANT1* polynucleotide encodes an *ANT1* polypeptide that retains substantially the same biological function or activity as the mature *ANT1* polypeptide encoded by the polynucleotide set forth as SEQ ID NO:1 (i.e. results in an *ANT1* phenotype when overexpressed in a plant).

[0068] Variants also include fragments of the *ANT1* polynucleotide of the invention, which can be used to synthesize a full-length *ANT1* polynucleotide. Preferred embodiments include polynucleotides encoding polypeptide variants wherein 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues of an *ANT1* polypeptide sequence of the invention are substituted, added or deleted, in any combination. Particularly preferred are substitutions, additions, and deletions that are silent such that they do not alter the properties or activities of the polynucleotide or polypeptide.

[0069] A nucleotide sequence encoding an *ANT1* polypeptide can also be used to construct hybridization probes for further genetic analysis. Screening of a cDNA or

genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook *et al.*, 1989). Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook *et al.*, *supra*.

[0070] The probes or portions thereof may also be employed in PCR techniques to generate a pool of sequences for identification of closely related *ANT1* sequences. When *ANT1* sequences are intended for use as probes, a particular portion of an *ANT1* encoding sequence, for example a highly conserved portion of the coding sequence may be used.

[0071] For example, an *ANT1* nucleotide sequence may be used as a hybridization probe for a cDNA library to isolate genes, for example, those encoding naturally-occurring variants of *ANT1* from other plant species, which have a desired level of sequence identity to the *ANT1* nucleotide sequence disclosed in SEQ ID NO:1. Exemplary probes have a length of about 20 to about 50 bases.

[0072] In another exemplary approach, a nucleic acid encoding an *ANT1* polypeptide may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein, and, if necessary, using conventional primer extension procedures as described in Sambrook *et al.*, *supra*, to detect *ANT1* precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

[0073] As discussed above, nucleic acid sequences of this invention may include genomic, cDNA or mRNA sequence. By "encoding" is meant that the sequence corresponds to a particular amino acid sequence either in a sense or anti-sense orientation. By "extrachromosomal" is meant that the sequence is outside of the plant genome of which it is naturally associated. By "recombinant" is meant that the sequence contains a genetically engineered modification through manipulation via mutagenesis, restriction enzymes, and the like.

[0074] Once the desired form of an *ANT1* nucleic acid sequence, homologue, variant or fragment thereof, is obtained, it may be modified in a variety of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, transversions, deletions, and insertions may be performed on the naturally occurring sequence.

[0075] With or without such modification, the desired form of the *ANT1* nucleic acid sequence, homologue, variant or fragment thereof, may be incorporated into a plant expression vector for transformation of plant cells.

***ANT1* Polypeptides**

- [0076] In one preferred embodiment, the invention provides an *ANT1* polypeptide, having a native mature or full-length *ANT1* polypeptide sequence comprising the sequence presented in SEQ ID NO:2. A *ANT1* polypeptide of the invention can be the mature *ANT1* polypeptide, part of a fusion protein or a fragment or variant of the *ANT1* polypeptide sequence presented in SEQ ID NO:2.
- [0077] Ordinarily, an *ANT1* polypeptide of the invention has at least 50% to 60% identity to an *ANT1* amino acid sequence over its entire length. More preferable are *ANT1* polypeptide sequences that comprise a region having at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the *ANT1* polypeptide sequence of SEQ ID NO:2.
- [0078] Fragments and variants of the *ANT1* polypeptide sequence of SEQ ID NO:2, are also considered to be a part of the invention. A fragment is a variant polypeptide that has an amino acid sequence that is entirely the same as part but not all of the amino acid sequence of the previously described polypeptides. Exemplary fragments comprises at least 10, 20, 30, 40, 50, 75, or 100 contiguous amino acids of SEQ ID NO:2. The fragments can be "free-standing" or comprised within a larger polypeptide of which the fragment forms a part or a region, most preferably as a single continuous region. Preferred fragments are biologically active fragments, which are those fragments that mediate activities of the polypeptides of the invention, including those with similar activity or improved activity or with a decreased activity. Also included are those fragments that antigenic or immunogenic in an animal, particularly a human.
- [0079] *ANT1* polypeptides of the invention also include polypeptides that vary from the *ANT1* polypeptide sequence of SEQ ID NO:2. These variants may be substitutional, insertional or deletional variants. The variants typically exhibit the same qualitative biological activity as the naturally occurring analogue, although variants can also be selected which have modified characteristics as further described below.
- [0080] A "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively.
- [0081] An "insertion" or "addition" is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to the naturally occurring sequence.
- [0082] A "deletion" is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.



- [0083] Amino acid substitutions are typically of single residues; insertions usually will be on the order of from about 1 to 20 amino acids, although considerably larger insertions may be tolerated. Deletions range from about 1 to about 20 residues, although in some cases deletions may be much larger.
- [0084] Substitutions, deletions, insertions or any combination thereof may be used to arrive at a final derivative. Generally these changes are done on a few amino acids to minimize the alteration of the molecule. However, larger changes may be tolerated in certain circumstances.
- [0085] Amino acid substitutions can be the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with a serine, *i.e.*, conservative amino acid replacements. Insertions or deletions may optionally be in the range of 1 to 5 amino acids.
- [0086] Substitutions are generally made in accordance with known "conservative substitutions". A "conservative substitution" refers to the substitution of an amino acid in one class by an amino acid in the same class, where a class is defined by common physicochemical amino acid side chain properties and high substitution frequencies in homologous proteins found in nature (as determined, *e.g.*, by a standard Dayhoff frequency exchange matrix or BLOSUM matrix). (*See generally*, Doolittle, R.F., 1986.)
- [0087] A "non-conservative substitution" refers to the substitution of an amino acid in one class with an amino acid from another class.
- [0088] *ANT1* polypeptide variants typically exhibit the same qualitative biological activity as the naturally occurring analogue, although variants also are selected to modify the characteristics of the *ANT1* polypeptide, as needed. For example, glycosylation sites, and more particularly one or more O-linked or N-linked glycosylation sites may be altered or removed. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the *ANT1* polypeptide, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.
- [0089] The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed mutagenesis [Carter *et al.*, 1986; Zoller *et al.*, 1987], cassette mutagenesis [Wells *et al.*, 1985], restriction selection mutagenesis [Wells *et al.*, 1986] or other known techniques can be performed on the cloned DNA to produce the *ANT1* polypeptide encoding variant DNA.

- [0090] Also included within the definition of *ANT1* polypeptides are other related *ANT1* polypeptides. Thus, probe or degenerate PCR primer sequences may be used to find other related polypeptides. Useful probe or primer sequences may be designed to all or part of the *ANT1* polypeptide sequence, or to sequences outside the coding region. As is generally known in the art, preferred PCR primers are from about 15 to about 35 nucleotides in length, with from about 20 to about 30 being preferred, and may contain inosine as needed. The conditions for the PCR reaction are generally known in the art.
- [0091] Covalent modifications of *ANT1* polypeptides are also included within the scope of this invention. For example, the invention provides *ANT1* polypeptides that are a mature protein and may comprise additional amino or carboxyl-terminal amino acids, or amino acids within the mature polypeptide (for example, when the mature form of the protein has more than one polypeptide chain). Such sequences can, for example, play a role in the processing of a protein from a precursor to a mature form, allow protein transport, shorten or lengthen protein half-life, or facilitate manipulation of the protein in assays or production. It is contemplated that cellular enzymes can be used to remove any additional amino acids from the mature protein. [See, *e.g.*, Creighton, TE, 1983].
- [0092] In a preferred embodiment, overexpression of an *ANT1* polypeptide or variant thereof is associated with the *ANT1* phenotype.

#### Antibodies

- [0093] The present invention further provides anti-*ANT1* polypeptide antibodies. The antibodies may be polyclonal, monoclonal, humanized, bispecific or heteroconjugate antibodies.
- [0094] Methods of preparing polyclonal antibodies are known to the skilled artisan. Such polyclonal antibodies can be produced in a mammal, for example, following one or more injections of an immunizing agent, and preferably, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected into the mammal by a series of subcutaneous or intraperitoneal injections. The immunizing agent may include an *ANT1* polypeptide or a fusion protein thereof. It may be useful to conjugate the antigen to a protein known to be immunogenic in the mammal being immunized. The immunization protocol may be determined by one skilled in the art based on standard protocols or by routine experimentation.

[0095] Alternatively, the anti-*ANT1* polypeptide antibodies may be monoclonal antibodies. Monoclonal antibodies may be produced by hybridomas, wherein a mouse, hamster, or other appropriate host animal, is immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent [Kohler *et al.*, 1975]. Monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567.

[0096] The anti-*ANT1* polypeptide antibodies of the invention may further comprise humanized antibodies or human antibodies. The term "humanized antibody" refers to humanized forms of non-human (*e.g.*, murine) antibodies that are chimeric antibodies, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub> or other antigen-binding partial sequences of antibodies) which contain some portion of the sequence derived from non-human antibody. Methods for humanizing non-human antibodies are well known in the art, as further detailed in Jones *et al.*, 1986; Riechmann *et al.*, 1988; and Verhoeven *et al.*, 1988. Methods for producing human antibodies are also known in the art. See, *e.g.*, Jakobovits, A, *et al.*, 1995; Jakobovits, A, 1995.

[0097] In one exemplary approach, anti-*ANT1* polyclonal antibodies are used for gene isolation. Western blot analysis may be conducted to determine that *ANT1* or a related protein is present in a crude extract of a particular plant species. When reactivity is observed, genes encoding the related protein may be isolated by screening expression libraries representing the particular plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.*, 1989.

#### **Utility Of the *ANT1* Phenotype and Gene**

[0098] From the foregoing, it can be appreciated that the *ANT1* nucleotide sequence, protein sequence and phenotype find utility in modulated expression of the *ANT1* protein and the development of non-native phenotypes associated with such modulated expression.

[0099] The *ANT1* phenotype has features which distinguish the mutant from wild type plants, including modified leaf color, modified flower color and modified fruit color.

[00100] Anthocyanins are known to contribute to leaf color, flower color and fruit color. Anthocyanins are a group of water-soluble flavonoids that impart pink to purple color to leaves and other organs (Harbone *et al.*, 1988). Anthocyanins have been associated

with many important physiological and developmental functions in the plants, including, but not limited to: (1) modification of the quantity and quality of captured light (Barker et al., 1977); (2) protection from the effects of UV-B radiation (Burger and Edwards, 1996 and Klaper et al., 1996); (3) defense against herbivores (Coley and Kusar, 1996); (4) protection from photoinhibition (Gould et al., 1995 and Dodd et al., 1998); and (5) scavenging of reactive oxygen intermediates in stressful environments (Furuta et al., 1995; Sherwin et al., 1998; and Yamasaki 1997). The anthocyanins have demonstrated anti-oxidant activity, suggesting a role in protecting against cancer, cardiovascular and liver diseases (Kamei et al., 1993; Suda et al., 1997; and Wang et al., 2000). See also the websites at [www.pslgroup.com/dg/39fb2.htm](http://www.pslgroup.com/dg/39fb2.htm), [www.wellweb.com/nutri/phytochemicals.htm](http://www.wellweb.com/nutri/phytochemicals.htm), and [www.nal.usda.gov/ttic/tektran/data/000007/19/0000071970.html](http://www.nal.usda.gov/ttic/tektran/data/000007/19/0000071970.html).

- [00101] Considering this, the *ANT1* phenotype described herein is not only colorful and hence finds utility in enhancement of the decorative value of ornamental plants, flowers and food, it also offers the potential for health benefits when the *ANT1* phenotype is expressed in plant varieties used as foods or food additives.
- [00102] In one aspect, the modified leaf, flower and fruit color of plants having the *ANT1* phenotype finds utility in the development of improved ornamental plants, fruits and/or cut flowers.
- [00103] In another aspect, the modified anthocyanin content in plants having the *ANT1* phenotype finds utility in plant-derived food and food additives.
- [00104] In another aspect, as further described in the Examples, the *ANT1* gene has utility as a transformation marker in genetically manipulated plants.
- [00105] In practicing the invention, the *ANT1* phenotype and modified *ANT1* expression is generally applicable to any type of plant, as further detailed below.
- [00106] The methods described herein are generally applicable to all plants. Although activation tagging and gene identification was carried out in tomato, following identification of a nucleic acid sequence and associated phenotype, the selected gene, a homologue, variant or fragment thereof, may be expressed in any type of plant. In one aspect, the invention is directed to fruit- and vegetable-bearing plants.
- [00107] The invention is generally applicable to plants which produce fleshy fruits; for example but not limited to, tomato (*Lycopersicum*); grape (*Vitis*); ; strawberry (*Fragaria*); raspberry, blackberry, loganberry (*Rubus*); currants and gooseberry (*Ribes*); blueberry, bilberry, whortleberry, cranberry (*Vaccinium*); kiwifruit and Chinese gooseberry (*Actinida*); apple

(*Malus*); pear (*Pyrus*); melons (*Cucumis sp.*) members of the *Prunus* genera, e.g. plum, cherry, nectarine and peach; sapota (*Manilkara zapotilla*); mango; avocado; apricot; peaches; cherries; pineapple; papaya; passion fruit; citrus; date palm; banana; plantain; and fig.

[00108] Similarly, the invention is applicable to vegetable plants, including, but not limited to sugar beets, green beans, broccoli, brussel sprouts, cabbage, celery, chard, cucumbers, eggplants, peppers, pumpkins, rhubarb, winter squash, summer squash, zucchini, lettuce, radish, carrot, pea, potato, corn, murraya and herbs.

[00109] In a related aspect, the invention is directed to the cut flower industry, grain-producing plants, oil-producing plants and nut-producing plants, as well as other crops including, but not limited to, cotton (*Gossypium*), alfalfa (*Medicago sativa*), flax (*Linum usitatissimum*), tobacco (*Nicotiana*), turfgrass (*Poaceae* family), and other forage crops.

[00110] References describing suitable transformation techniques for these and other plants are listed in Patent Application Serial No. 09/846,758.

[00111] The skilled artisan will recognize that a wide variety of transformation techniques exist in the art, and new techniques are continually becoming available. Any technique that is suitable for the target host plant can be employed within the scope of the present invention. For example, the constructs can be introduced in a variety of forms including, but not limited to as a strand of DNA, in a plasmid, or in an artificial chromosome. The introduction of the constructs into the target plant cells can be accomplished by a variety of techniques, including, but not limited to *Agrobacterium*-mediated transformation, electroporation, microinjection, microprojectile bombardment calcium-phosphate-DNA co-precipitation or liposome-mediated transformation of a heterologous nucleic acid construct comprising the *ANT1* coding sequence. The transformation of the plant is preferably permanent, *i.e.* by integration of the introduced expression constructs into the host plant genome, so that the introduced constructs are passed onto successive plant generations.

[00112] In one embodiment, binary Ti-based vector systems may be used to transfer and confirm the association between enhanced expression of an identified gene with a particular plant trait or phenotype. Standard *Agrobacterium* binary vectors are known to those of skill in the art and many are commercially available, such as pBI121 (Clontech Laboratories, Palo Alto, CA).

[00113] The optimal procedure for transformation of plants with *Agrobacterium* vectors will vary with the type of plant being transformed. Exemplary methods for

*Agrobacterium*-mediated transformation include transformation of explants of hypocotyl, shoot tip, stem or leaf tissue, derived from sterile seedlings and/or plantlets. Such transformed plants may be reproduced sexually, or by cell or tissue culture. *Agrobacterium* transformation has been previously described for a large number of different types of plants and methods for such transformation may be found in the scientific literature.

- [00114] Depending upon the intended use, a heterologous nucleic acid construct may be made which comprises a nucleic acid sequence associated with the *ANTI* phenotype, and which encodes the entire protein, or a biologically active portion thereof for transformation of plant cells and generation of transgenic plants.
- [00115] The expression of an *ANTI* nucleic acid sequence or a homologue, variant or fragment thereof may be carried out under the control of a constitutive, inducible or regulatable promoter. In some cases expression of the *ANTI* nucleic acid sequence or homologue, variant or fragment thereof may regulated in a developmental stage or tissue-associated or tissue-specific manner. Accordingly, expression of the nucleic acid coding sequences described herein may be regulated with respect to the level of expression, the tissue type(s) where expression takes place and/or developmental stage of expression leading to a wide spectrum of applications wherein the expression of an *ANTI* coding sequence is modulated in a plant.
- [00116] Strong promoters with enhancers may result in a high level of expression. When a low level of basal activity is desired, a weak promoter may be a better choice. Expression of *ANTI* nucleic acid sequence or homologue, variant or fragment thereof may also be controlled at the level of transcription, by the use of cell type specific promoters or promoter elements in the plant expression vector.
- [00117] Numerous promoters useful for heterologous gene expression are available. Exemplary constitutive promoters include the raspberry E4 promoter (U.S. Patent Nos. 5,783,393 and 5,783,394), the 35S CaMV (Jones JD *et al.*, 1992), the CsVMV promoter (Verdaguer B *et al.*, 1998) and the melon actin promoter. Exemplary tissue-specific promoters include the tomato E4 and E8 promoters (U.S. Patent No. 5,859,330) and the tomato 2AII gene promoter (Van Haaren MJJ *et al.*, 1993).
- [00118] When *ANTI* sequences are intended for use as probes, a particular portion of an *ANTI* encoding sequence, for example a highly conserved portion of a coding sequence may be used.

**[00119]** In yet another aspect, in some cases it may be desirable to inhibit the expression of endogenous *ANT1* sequences in a host cell. Exemplary methods for practicing this aspect of the invention include, but are not limited to antisense suppression (Smith, *et al.*, 1988); co-suppression (Napoli, *et al.*, 1989); ribozymes (PCT Publication WO 97/10328); and combinations of sense and antisense (Waterhouse, *et al.*, 1998). Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence. In some cases, it may be desirable to inhibit expression of the *ANT1* nucleotide sequence. This may be accomplished using procedures generally employed by those of skill in the art together with the *ANT1* nucleotide sequence provided herein.

**[00120]** Standard molecular and genetic tests may be performed to analyze the association between a cloned gene and an observed phenotype. A number of other techniques that are useful for determining (predicting or confirming) the function of a gene or gene product in plants are described below.

#### **DNA/RNA analysis**

**[00121]** DNA taken from a mutant plant may be sequenced to identify the mutation at the nucleotide level. The mutant phenotype may be rescued by overexpressing the wild type (WT) gene. The stage- and tissue-specific gene expression patterns in mutant vs. WT lines, for instance, by *in situ* hybridization, may be determined. Analysis of the methylation status of the gene, especially flanking regulatory regions, may be performed. Other suitable techniques include overexpression, ectopic expression, expression in other plant species and gene knock-out (reverse genetics, targeted knock-out, viral induced gene silencing (VIGS, see Baulcombe D, 1999).

**[00122]** In a preferred application, microarray analysis, also known as expression profiling or transcript profiling, is used to simultaneously measure differences or induced changes in the expression of many different genes. Techniques for microarray analysis are well known in the art (Schena M *et al.*, Science (1995) 270:467-470; Baldwin D *et al.*, 1999; Dangond F, Physiol Genomics (2000) 2:53-58; van Hal NL *et al.*, J Biotechnol (2000) 78:271-280; Richmond T and Somerville S, Curr Opin Plant Biol (2000) 3:108-116). Microarray analysis of individual tagged lines may be carried out, especially those from which genes have been isolated. Such analysis can identify

other genes that are coordinately regulated as a consequence of the overexpression of the gene of interest, which may help to place an unknown gene in a particular pathway.

#### **Gene Product Analysis**

- [00123] Analysis of gene products may include recombinant protein expression, antisera production, immunolocalization, biochemical assays for catalytic or other activity, analysis of phosphorylation status, and analysis of interaction with other proteins via yeast two-hybrid assays.

#### **Pathway Analysis**

- [00124] Pathway analysis may include placing a gene or gene product within a particular biochemical or signaling pathway based on its overexpression phenotype or by sequence homology with related genes. Alternatively, analysis may comprise genetic crosses with WT lines and other mutant lines (creating double mutants) to order the gene in a pathway, or determining the effect of a mutation on expression of downstream "reporter" genes in a pathway.

#### **Other Analyses**

- [00125] Other analyses may be performed to determine or confirm the participation of the isolated gene and its product in a particular metabolic or signaling pathway, and to help determine gene function.
- [00126] All publications, patents and patent applications are herein expressly incorporated by reference in their entirety.
- [00127] While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention.

### **EXAMPLE 1**

#### **Generation of Plants with an *ANTI* Phenotype by Transformation with an Activation Tagging Construct**

##### **I. *Agrobacterium* vector preparation.**

- [00128] Mutants were generated using a modified version of the activation tagging "ACTTAG" vector, pSKI015 (GenBank Identifier [GI] 6537289; Weigel D *et al.*, 2000). This binary vector, called pAG3202, contains the following components: the



pSKI backbone; a 4X 35S enhancer consisting of four tandem repeats of the enhancer region from the CaMV 35S promoter including 4 AluI-EcoRV fragments in tandem, 129 bp of CaMV sequence associated with each tandem AluI-EcoRV repeat, and an additional 7 bp repeated sequence that is not in the 35S enhancer region of the native CaMV genome; the *nptII* selectable marker under the control of a raspberry E4 (RE4) promoter; an *Agrobacterium* gene 7 termination element located downstream of the *nptII* gene, adjacent the left border of the plasmid. The pAG3202 sequence is provided in SEQ ID NO: 3.

[00129] Single colonies of *Agrobacterium tumefaciens* strains EHA 105/EHA 101/GV3101 containing the binary plasmid pAG3202 were grown in MGL medium at pH 5.4 overnight and diluted to approximately  $5 \times 10^8$  cells/ml with MGL or liquid plant co-cultivation medium.

[00130] For long-term storage, PCR-positive colonies were grown in selective media, glycerol added to a final concentration of 30% and cultures quick frozen, then stored at  $-80^\circ\text{C}$ . For the initiation of dense *Agrobacterium* cultures for plant transformation, stock cultures were grown in selective media, glycerol added to a final concentration of 30%, and a number of 20  $\mu\text{l}$  aliquots quick frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$ .

## II. Transformation and Selection of Micro-Tom Mutants

[00131] Activation tagging mutants were generated in tomato cv. Micro-Tom using *Agrobacterium*-mediated transformation. Sterile seedlings and plantlets were used as the source of explants. More specifically hypocotyl tissue was transformed.

[00132] Seeds of (*Lycopersium esculentum*) were surface sterilized in 25% bleach with tween-20 for 15 minutes and rinsed with sterile water before plating on seed germination medium (MS salts, Nitsch vitamins, 3% sucrose and 0.7% agar, pH 5.8), modified by the addition of auxin and/or cytokinins and giberellic acid as necessary. The cultures were incubated at  $24^\circ\text{C}$  with a 16 hr photo period ( $50\text{--}60 \mu\text{mol.m}^{-2}\text{s}^{-1}$ ). Seven to ten day old seedlings and one month old *in vitro* plants were used for hypocotyl explants.

[00133] Hypocotyls were cut into 3-5 mm segments, then immersed in bacterial suspension, blotted on sterile filter paper and placed on co-cultivation medium. The explants were immersed in bacterial suspension, blotted on sterile filter paper and placed on co-cultivation medium (MS salts, LS vitamins, 3% sucrose, 0.1 mg/l kinetin,

0.2 mg/l 2,4-D, 200 mg/l potassium acid phosphate, 50  $\mu$ M acetosyringone and 0.7% agar, pH 5.4) for 2-3 days.

[00134] After two to three days of co-culture, the explants were transferred to shoot regeneration medium containing MS salts, Nitsch vitamins, 3% sucrose, 2 mg/l zeatin, 500 mg/l carbenicillin, 200mg/L timetin and 0.7% agar at pH 5.8, supplemented with the antibiotic, kanamycin at 75 – 400 mg/l in order to select for *nptII*-expressing transformants. The selection level of antibiotic was gradually raised over an 8-week period based on the tissue response.

[00135] The explants were transferred to fresh medium every two weeks. Initiation of callus with signs of shoot initials was observed from 3-6 weeks depending on the type of explant. Callusing and shoot regeneration was observed to continue over approximately 4 months after which the explant tissues decline. A purple callus was observed among the tissue growing on the selection medium. Regenerated shoots displayed a variety of color phenotypes and were entirely green, entirely purple, or mix of green and purple to various degrees. Green shoots of approximately 1 cm in size with distinct shoot meristems were excised from the callus and transferred to root induction medium containing MS salts, Nitsch vitamins, 3% sucrose, 1 mg/l IBA, 50 mg/l kanamycin, 100 mg/l carbenicillin or 100mg/L timetin and 0.7% agar, pH 5.8. The rooted plants were out-planted to soil in a Biosafety greenhouse.

[00136] Plants were transported to greenhouse facilities, potted up in 3.5" pots tagged for plant identification.

[00137] Transformants were observed at the callus stage and after T<sub>1</sub> plants were established in the greenhouse for phenotypic variations relative to wild-type Micro-Tom plants. To achieve this, several wild-type plants were kept in close proximity to the transgenic plants. Each plant was observed closely twice a week with observations noted and documented by photographs.

[00138] Images of each pool of 8 plants were recorded using a Digital camera (DC-260), and morphology observations were made at about four weeks after planting.

[00139] Eleven Micro-Tom lines were developed from the callus originally identified by its purple color and purple shoots at the caulogenic stage. The clonal plant lines were identified as having modified leaf color with a heavy purple cast on leaves, modified flower color characterized by purple striations on petals and sepals and flowers with a purple cast mixed with the normal yellow color of the corolla. The plants were also observed to exhibit a modified fruit color described as a deeper red color relative to

wild type Micro-Tom plants. The clonal plant lines (mutants) were designated *Anthocyanin1* ("ANT1").

[00140] The *ANT1* mutant was identified from fewer than 2000 individual Micro-Tom tomato ACTTAG lines that were developed following tissue culture transformation with the binary plasmid pAG3202, and selection on kanamycin-containing medium.

[00141] Observations were made and photos taken of the clonal T<sub>1</sub> *ANT1* plant lines that exhibited the *ANT1* phenotype, designated H000001484, H000001624, H000001708, H000001709, H000001710, H000001711, H000001712, H000001713, H000001715, H000001716 and H000001717.

[00142] Seeds were collected from T<sub>1</sub> plants from line H000001624 and grown to generate T<sub>2</sub> plants. From the 11 out of the 18 seeds that germinated, and 8 plants displayed purple coloration, confirming that *ANT1* is a dominant mutation.

[00143] The results indicated that *ANT1* is a gain of function trait, expected from activation tagging based over-expression of a native gene.

## **EXAMPLE 2**

### **Characterization of Plants That Exhibit the *ANT1* Phenotype**

[00143] Micro-Tom genomic DNA was extracted from the H000001484 clone of the activation tagged mutant originally identified at the callus stage, in sufficient yield and quality for plasmid rescue of activation tagged plant lines using the procedure described below. Further analysis was performed using combined tissue derived from the H000001624, H000001708, H000001709, H000001710, H000001711, H000001712, H000001713, H000001715, H000001716 and H000001717 plant lines.

#### **I. Micro-Tom Tomato Genomic DNA Extraction**

[00144] Nucleon™ PhytoPure™ systems (Plant and fungal DNA extraction kits) from Amersham™ were used for extracting genomic DNA. Methods were essentially as follows:

[00145] 1.0g of fresh tissue from the H000001484 clone was ground in liquid nitrogen to yield a free flowing powder, then transferred to a 15 ml polypropylene centrifuge tube. 4.6 ml of Reagent 1 from the Nucleon Phytopure kit was added with thorough mixing, followed by addition of 1.5 ml of Reagent 2 from the Nucleon Phytopure kit, with inversion until a homogeneous mixture was obtained. The mixture was incubated at 65°C in a shaking water bath for 10 minutes, and placed on ice for 20 minutes. The

samples were removed from the ice, 2 ml of  $-20^{\circ}\text{C}$  chloroform added, mixed and centrifuged at 1300g for 10 minutes. The supernatant was transferred into a fresh tube, 2 ml cold chloroform, 200  $\mu\text{l}$  of Nucleon PhytoPure DNA extraction resin suspension added and the mixture shaken on a tilt shaker for 10 minutes at room temperature, then centrifuged at 1300g for 10 minutes. Without disturbing the Nucleon resin suspension layer, the upper DNA-containing phase was transferred into a fresh tube, centrifuged at 9500 rpm for 30 minutes to clarify the transferred aqueous phase if the upper phase appeared cloudy, an equal volume of cold isopropanol added, and the tube gently inverted until DNA precipitated. It was then pelleted by centrifugation, washed with cold 70% ethanol, pelleted again, and air-dried.

[00146] DNA was resuspended in TE buffer (10 mM Tris. HCl, pH 7.4, 1 mM EDTA), containing RNase, incubated at  $55^{\circ}\text{C}$  for 15 minutes, further extracted phenol/chloroform, then chloroform, run on a 1% agarose gel to check the DNA Quality, the DNA concentration determined by a DNA fluorometer (Hoeffer DyNA Quant 200).

[00147] DNA extracted from shoots of the H000001484 *ANT1* clone at the caulogenic callus stage and from wild type plants was PCR-amplified using primers that amplify a 35S enhancer sequence, and primers that amplify a region of the pBluescript vector sequence in pAG3202. Amplification using primers that span the 35S enhancer region resulted in a ladder of products, indicating that all four copies of the 35S enhancer were present. Amplification using primers to the pBluescript vector was done primarily to detect the T-DNA insert(s) in transformed plants and has been optimized for the following conditions: annealing temp:  $57^{\circ}\text{C}$ , 30 cycles [ $94^{\circ}\text{C}$ , 30sec;  $57^{\circ}\text{C}$ , 1 min;  $72^{\circ}\text{C}$ , 1 min] 1 cycle [ $72^{\circ}\text{C}$ , 7 min].

[00148] The ACTTAG<sup>TM</sup> line, H000001484 (*ANT1*), was confirmed as positive for the presence of 35S enhancer and pAG3202 vector sequences by PCR, and as positive for Southern hybridization verifying genomic integration of the ACTTAG DNA and showing the presence of a single T-DNA insertion in the clonal transgenic line.

## II. Plasmid Rescue

[00149] Genomic DNA from the H000001484 clonal line was digested by the restriction enzymes used in Southern Hybridization. The restriction fragments were self-ligated and used to transform the *E. coli* cells. The plasmids that contained a full-length

pBluescript vector, 4X 35S enhancer, and a right border T-DNA flanking genomic DNA fragment were rescued.

[00150] More specifically, genomic DNA was digested with Hind III and Xho I under standard reaction conditions at 37°C overnight.

[00151] The ligation reactions were set up containing the following and left at 16°C overnight:

Digested Genomic DNA	40 µl
5X Ligation Buffer	50 µl
Ligase (Gibcol, 1U/µl)	10 µl
ddH <sub>2</sub> O	150 µl

[00152] The ligated DNA precipitated, resuspended in ddH<sub>2</sub>O and used to transform *E. coli* SURE cells (Stratagene) via electroporation, with 10 pg of pUC18 plasmid as a control.

[00153] The transformation mixture was spread on two LB-plates containing 100 µg/ml ampicillin and incubated overnight at 37°C. Single colonies were picked from the plates and used to start a 5 ml LB-ampicillin broth culture from each colony by culturing overnight at 37°C. The plasmid was extracted from the culture and restriction digested to confirm the size of genomic insertion.

### III. Sequencing Of Rescued Plasmids

[00154] Sequencing was accomplished using a ABI Prism BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystem), AmpliTaq DNA Polymerase (Perkin-Elmer), an ABI Prism™ 310 Genetic Analyzer (Perkin-Elmer) and sequence analysis software, *e.g.*, Sequencer™ 3.1.1 or MacVector 6.5.3. Sequencing was done essentially according to manufacturers' protocols

[00155] The left ends of plasmids rescued were sequenced across the right T-DNA border.

[00156] The rescued sequence was subjected to analysis using the BLAST sequence comparison programs at the [www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST) website. A basic BLASTN search identified a sequence with 31% identity to the Anthocyanin 2 (An2) mRNA of *Petunia integrifolia* (GI 7673087 and 7673085). The presence of an open reading frame (*i.e.*, the *ANT1* cDNA) was predicted using the BLASTX program.

[00157] RT-PCR analysis confirmed that the gene whose nucleotide sequence is presented as SEQ ID NO:1 (*ANT1*) was specifically overexpressed in tissue from plants having the *ANT1* phenotype. Specifically, RNA was extracted from combined tissues derived from the H000001624 clonal plant line, which exhibited the *ANT1* phenotype, and from wild type plants. RT-PCR was performed using primers specific to the sequence presented as SEQ ID NO:1 and a constitutively expressed actin gene (positive control). The results showed that plants displaying the *ANT1* phenotype over-expressed the mRNA for the *ANT1* gene, indicating the enhanced expression of the *ANT1* gene correlated with the *ANT1* phenotype.

[00158] The amino acid sequence predicted from the *ANT1* nucleic acid sequence was determined using Vector NTI (InforMax, North Bethesda, MD) and is presented in SEQ ID NO:2. A Basic BLASTP 2.0.11 search using the [ncbi.nlm.nih.gov/BLAST](http://ncbi.nlm.nih.gov/BLAST) website and the predicted *ANT1* amino acid sequence was conducted. Results indicated that the predicted *ANT1* protein sequence has 49% identity to the *Petunia integrifolia* An2 protein sequence (GI 7673088 and 7673086) and 65%-85% identity to several *Myb*-related transcription factors in the N-terminal region, from approximately aa 1-120 of SEQ ID NO:2. These *Myb*-related proteins included An2 from *Petunia x hybrida* (GI 7673084), the *Zea mays* C1-I (GI 22214), the *Zea mays* PL transcription factor (GI 2343273) and an *Arabidopsis* transcription factor (GI 3941508). The *Petunia* An2 gene is a regulator of the Anthocyanin biosynthetic pathway (Quattrocchio et al, 1999).

[00159] These results suggest that *ANT1* is associated with modified leaf, flower or fruit color in Micro-tomato.

### **EXAMPLE 3**

#### **Confirmation of Phenotype/Genotype Association in Micro-tomato**

[00160] In order to further confirm the association between the *ANT1* phenotype and the *ANT1* gene presented in SEQ ID NO: 1, a genomic fragment comprising the *ANT1* gene, provided in SEQ ID NO:4, was over-expressed in wild type Micro-Tom plants. Specifically, this 1012 bp genomic fragment, including the *ANT1* coding regions, was cloned into the multiple cloning site (MCS) of the binary vector pAG2370. pAG2370, whose sequence is provided in SEQ ID NO:5, comprises the vector backbone from the binary vector pBIN19 (GI:256363), T-DNA left and right border fragments, and, between border fragments, the CsVMV promoter sequence and a Nos termination

sequence for controlling expression of the inserted gene, and the neomycin phosphotransferase (NPTII) gene, which confers kanamycin resistance, whose expression is controlled by the RE4 promoter (US Patent No. 6054635) and the G7 termination sequence. The *ANT1* fragment was cloned into *Sma*I/*Spe*I sites of pAG2370, inserted between the CsVMV promoter region, proximal to the 5' end of genomic fragment, and the Nos termination sequence, proximal to the 3' end of the genomic fragment. The pAG2370-*ANT1* construct was transformed into *Agrobacterium tumefaciens* by electroporation.

[00161] The pAG2370-*ANT1* construct described above was introduced into wild-type Micro-Tom plants via *Agrobacterium*-mediated transformation, essentially as described in Example 1. Briefly, explants were dissected from Micro-Tom seedlings. Explants were inoculated by soaking in the *Agrobacterium* suspension for 15 to 120 minutes, blotted on sterile filter paper to remove excess bacteria, and plated. Explants were co-cultivated in non-selective media for 2-4 days at 24°C with a 16-hour photoperiod, after which they were transferred to selective media (with kanamycin) and returned to the growth room. Explants were transferred to fresh medium every two weeks until shoots were 0.5 to 1 cm tall. Shoots were excised from the explants, placed on selective medium with kanamycin in Phytatrays (Sigma), and returned to the growth room for two to four weeks. Shoots were observed for rooting, and rooted shoots were out-planted to soil and acclimated to the greenhouse. The transformation process generated 64 independent T<sub>0</sub> events. Morphological observations demonstrated that 45 transgenic plants displayed the *ANT1* purple color phenotype and were either partially or entirely purple. Tissue was collected from six T<sub>1</sub> plants showing the *ANT1* phenotype, and RT-PCR was carried out using wild type as a control. While no *ANT1* gene expression could be detected in the wild-type control, five out of the six plants displaying the *ANT1* phenotype over-expressed the *ANT1* transcript. The internal control experiments, using a constitutively expressed actin gene, showed that all samples had similar levels of the actin expression.

#### **EXAMPLE 4**

##### **Confirmation of Phenotype/Genotype Association in *Arabidopsis***

[00163] In order to further confirm the association between the *ANT1* phenotype and the *ANT1* gene in plants other than Micro-Tom, the *ANT1* gene was introduced into and over-expressed in wild type *Arabidopsis thaliana*.

[00164] The pAG2370-*ANT1* construct described above was introduced into wild-type *Arabidopsis* plants via *Agrobacterium*-mediated transformation using standard vacuum infiltration methods. All infiltrated seeds were plated in selective media containing kanamycin, and kanamycin-resistant T<sub>1</sub> plants were transplanted to 72-cell flats. The transformation process generated 10 independent T<sub>0</sub> events, of which seven displayed the *ANT1* purple coloration phenotype in at least part of the plant. Tissue was collected from four T<sub>1</sub> plants showing the *ANT1* phenotype, and RT-PCR was carried out using wild type as a control. While no *ANT1* gene expression could be detected in the wild-type control, all plants displaying the *ANT1* phenotype over-expressed the *ANT1* transcript. The internal control experiments, using a constitutively expressed actin gene, showed that all samples had similar levels of the actin expression.

### **EXAMPLE 5**

#### **Confirmation of Phenotype/Genotype Association in Tobacco**

[00166] In order to further confirm the association between the *ANT1* phenotype and the *ANT1* gene in plants other than Micro-Tom, the *ANT1* gene was introduced into and over-expressed in wild type *Nicotiana tabacum* (tobacco, Wisconsin-38 type).

[00167] The pAG2370-*ANT1* construct described above was introduced into wild-type tobacco plants via *Agrobacterium*-mediated transformation using essentially the following methods. In order to generate tobacco plants for transformation, tobacco seeds were germinated as follows: seeds were shaken about ten minutes on a lab shaker, in a solution containing approximately 1.3% to 2.1% sodium hypochlorite and one drop of Tween-20 (Polyoxyethylenesorbitan monolaurate) per 100 milliliters. Seeds were then washed in sterile water and sterilely transferred to the surface of TbSG medium (4.3 g/l Murashige and Skoog salts, Phytotech; 1 ml/l MS vitamins, Sigma; 30 g/l sucrose; 8 g/l agar, Sigma; pH adjusted to ~5.8) in petri dishes or Phytatrays (Sigma), 10-50 seeds per vessel, and incubated in light at 25°C. Tobacco plants were dissected on sterile filter paper moistened with sterile, deionized water or liquid TbCo medium (4.3 g/l Murashige and Skoog salts, Phytotech; 1 ml/l MS vitamins, Sigma; 30 g/l sucrose; 200 mg/l KH<sub>2</sub>PO<sub>4</sub>; 2 mg/l Indole-3-acetic acid; 0.25 mg/l Kinetin; 0 to 100 µM Acetosyringone; 7 g/l Agar, Sigma; pH adjusted to 5.4-5.6). Explants with cut edges on all sides could be generated by cutting the leaf from the plant, dissecting out and discarding the midvein, and cutting the leaf lamina into 3 to 5



mm squares. Alternatively, discs could be cut from the lamina using a sterilized cork borer.

[00168] Explants were inoculated by soaking for 15-120 minutes in *Agrobacterium* suspension (OD<sub>600</sub> between 0.175 and 0.225) prepared with the pAG2370-*ANTI* construct, then blotted and plated on TbCo medium. Explants were co-cultivated 2-4 days at 24°C with a 16-hour photoperiod, and then transferred to Tb selective medium (4.3 g/l Murashige and Skoog salts; 1 ml/l Nitsch and Nitsch vitamins, Duchefa; 30 g/l sucrose; 0.5 to 2 mg/l 6-Benzylaminopurine; 0 to 1 mg/l Naphthylacetic Acid; 0 to 750 mg/l Carbenicillin; 0 to 300 mg/l Timentin; 0 to 500 mg/l Kanamycin; 7 to 8 g/l Agar, Sigma; pH adjusted to ~5.8) containing kanamycin and re-transferred every two weeks until shoots were 0.5 to 1 cm tall. Shoots were excised from the explants, placed on TbR medium (4.3 g/l Murashige and Skoog salts; 1 ml/l Nitsch and Nitsch vitamins, Duchefa; 30 g/l sucrose; 0 to 1 mg/l Indole-3-butyric acid; 0 to 1 mg/l Naphthylacetic Acid; 0 to 100 mg/l Carbenicillin; 0 to 200 mg/l Timentin; 0 to 100 mg/l Kanamycin; 7 to 8 g/l Agar, Sigma; pH adjusted to ~5.8.) with kanamycin in Phytatrays, and grown two to four weeks, after which time the rooted shoots were planted to soil.

[00169] The transformation process generated 89 independent T<sub>0</sub> events, of which 54 displayed the *ANTI* purple coloration phenotype in at least part of the plant. Tissue was collected from five T<sub>1</sub> plants showing the *ANTI* phenotype, and RT-PCR was carried out using wild type as a control. While no *ANTI* gene expression could be detected in the wild-type control, all plants displaying the *ANTI* phenotype over-expressed the *ANTI* transcript. The internal control experiments, using a constitutively expressed actin gene, showed that all samples had similar levels of the actin expression.

#### **EXAMPLE 6**

##### **Use of the *ANTI* gene as a transformation marker in tomato and tobacco**

[00171] Having successfully recapitulated the *ANTI* phenotype in tomato and tobacco, as described above, we tested the utility of the *ANTI* gene for utility as a transformation marker, based on its characteristic purple color, in these species. We transformed tobacco and Micro-Tom explants with the pAG2370-*ANTI* vector, using methods described in the above Examples, grew the explants in the presence and absence of antibiotic (kanamycin), and compared transformation frequency based on rooting in

the presence of antibiotic in the media to transformation frequency based on purple color. Results are shown in the Table below.

[00172] Table 1:

Transformation frequency of tobacco and tomato, based on antibiotic selection or color

Species	Kanamycin in media	# explants	Transformation frequency based on rooting in presence of antibiotic	Transformation frequency based on purple color
<u>Tobacco</u> Wisconsin	+	82	126 % *	80%
	-	60	-	45%
<u>Tomato</u> Micro-Tom	+	103	77%	54%
	-	52	-	6%

\*This number reflects multiple transgenic events per original explant. When callus initiation occurs at two or three distinct points on the original explant, each is dissected and tested for shoot regeneration.

[00173] The results indicated that the *ANT1* gene could be successfully used for screening of positive transformants in cultures of tomato and tobacco, and may be useful in other plants as well.

## REFERENCES

- [00174] Altschul, S.F. *et al.*, *J. Mol. Biol.* 215:403-410, 1990.
- Altschul, S.F. *et al.*, *Nucleic Acids Res.* 25:3389-3402, 1997.
- Ausubel FM *et al.* Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., 1993.
- Baldwin D *et al.*, *Cur Opin Plant Biol.* 2(2):96-103, 1999.
- Barker DH *et al.*, *Plant Cell and Environment* 20: 617-624, 1977.
- Baulcombe D, *Arch Virol Suppl* 15:189-201, 1999.
- Behringer and Medford, *Plant Mol. Biol. Rep.* 10(2):190-198, 1992.
- Burger J and Edwards GE. *Plant and Cell Physiology* 37: 395-399, 1996.
- Carter *et al.*, *Nucl. Acids Res.* 13:4331, 1986.
- Christensen S *et al.*, 9<sup>th</sup> International Conference on *Arabidopsis* Research. Univ. of Wisconsin-Madison, June 24-28, 1998. Abstract 165.

- Coley PD and Kusar TA. In: Mulkey SS, Chazdon RL, Smith AP, eds. **TROPICAL FOREST PLANT ECOPHYSIOLOGY**. New York: Chapman and Hall 305-335, 1996.
- Cough, SJ and Bent, AF, *the Plant Journal* 16(6): 735-743, 1998.
- Crameri A and Stemmer WP, *Bio Techniques* 18(2):194-6, 1995.
- Creighton, T.E., PROTEINS: STRUCTURE AND MOLECULAR PROPERTIES, W.H. Freeman & Co., San Francisco, pp. 79-86, 1983.
- Dodd IC et al., *Journal of Experimental Botany* 49: 1437-1445, 1998.
- Doolittle, R.F., *OF URFS and ORFS* (University Science Books, CA, 1986.)
- Dooner et al, 1991, *Ann.Rev.Genet.* 25:179-199.
- Fang G et al., *Plant Cell.*, 1(1):141-50, 1989.
- Feldman et al., *Science* 243: 1351-1354, 1989.
- Fridborg I et al., *Plant Cell* 11: 1019-1032, 1999.
- Furuta S et al., *Sweetpotato Res Front* (KNAES, Japan) 1:3, 1995.
- Geest AH and Hall TC, *Plant Mol Biol* 32(4):579-88, 1996).
- Gelvin, S. B., Schilperoort, R. A., Varma, D. P. S., eds. **Plant Molecular Biology Manual** 1990.
- Glick, BR and Thompson, JE, Eds. **METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY**, p. 213-221, CRC Press, 1993.
- Gould KS, et al., *Nature* 378: 241-242, 1995.
- Harling et al., *EMBO J.* 16: 5855-66, 1997.
- Haught C et al. *BioTechniques* 16(1):47-48, 1994.
- Hayashi H et al., *Science* 258: 1350-1353, 1992.
- Harbone JB, **The Flavonoids: Recent Advances**. In: Goodwin TW, ed. **PLANT PIGMENTS**. London: Academic Press, 299-343, 1988.
- Jakobovits, A, et al., *Ann N Y Acad Sci* 764:525-35, 1995.
- Jakobovits, A, *Curr Opin Biotechnol* 6(5):561-6, 1995.
- Jayaram and Peterson, 1990, *Plant Breeding Reviews* 2:91-137; Coe, 1994, In 'The Maize Handbook', Freeling and Walbot, eds. Springer Verlag New York Inc., p. 279-281
- Jensen, L.G., et al., *Proc. Natl. Acad. Sci. USA* 93:3487-3491, 1996.
- Jones et al., *Nature* 321:522-525, 1986.
- Jones JD et al, *Transgenic Res* 1:285-297 1992.
- Kakimoto, *Science* 274: 982-5, 1996.
- Kamei H et al., *J Clin Exp Med* 164: 829, 1993.

- Kardailsky I *et al.*, *Science* 286: 1962-1965, 1999.
- Klaper R *et al.*, *Photochemistry and Photobiology* 63: 811-813, 1996.
- Kohler and Milstein, *Nature* 256:495, 1975.
- Kunkel TA *et al.*, *Methods Enzymol.* 204:125-39, 1991.
- Liu *et al.*, *Plant Journal* 8(3) 457-463, 1995.
- Marks and Feldman, *Plant Cell* 1:1053-1050, 1989.
- Meissner *et al.*, *The Plant Journal* 12(6) 1465-1472, 1997.
- Miklashevichs *et al.*, *Plant J.* 12: 489-98, 1997.
- Napoli, *et al.*, *Plant Cell* 2:279-289, 1990.
- Novak, J and Novak, L, *Promega Notes Magazine* Number 61:27, 1997.
- Omirulleh *et al.*, *Plant Mol Biol.* 21(3):415-28, 1993.
- Quattrocchio F *et al.*, 1999, *Plant Cell* 11:1433-1444
- Riechmann *et al.*, *Nature* 332:323-327, 1988.
- Sambrook *et al.* *Molecular Cloning: A Laboratory Manual (Second Edition)*, Cold Spring Harbor Press, Plainview, N.Y., 1989.
- Schaffer R, *et al.*, *Cell* 93: 1219-1229, 1998.
- Schell *et al.*, *Trends Plant Sci.* 3: 130, 1998.
- Scott, JW and Harbaugh, BK, University of Fla. Circular S-370, Dec. 1989.
- Sherwin HW and Farrant JM., *Plant Growth Regulation* 24: 203-210, 1998.
- Smith, *et al.*, *Nature* 334:724-726, 1988.
- Suda I, *et al.*, 1997. *Sweetpotato Res Front (KNAES, Japan)* 4:3, 1997.
- Van Haaren MJJ *et al.*, *Plant Mol Bio* 21:625-640, 1993.
- Verdaguer B *et al.*, *Plant Mol Biol* 37:1055-1067, 1998.
- Verhoeven *et al.*, *Science* 239:1534-1536, 1988.
- Walden *et. al.*, *EMBO J.* 13: 4729-36, 1994.
- Walden *et al.*, *Plant Mol. Biol.* 26: 1521-8, 1994.
- Wang CJ, *et al.*, *H Food Chem Toxicology* 38: 411-416, 2000.
- Waterhouse, *et al.*, *Proc. Natl. Acad. Sci. USA* 95:13959-13964, 1998.
- Wells *et al.*, *Gene* 34:315, 1985.
- Wells *et al.*, *Philos. Trans. R. Soc. London SerA* 317:415, 1986.
- Weigel D, *et al.*, *Plant Physiology*, 122:1003-1013, 2000.
- Wilson K *et al.*, *Plant Cell* 8: 659-671, 1996.
- Xu YL, *et al.*, *Plant Cell*, 11: 927-36, 1999.
- Yamasaki H *Trends in Plant Science* 2: 7-8, 1997.
- Zoller *et al.*, *Nucl. Acids Res.* 10:6487, 1987.

**IT IS CLAIMED:**

1. An isolated polynucleotide comprising a nucleic acid sequence which encodes or is complementary to a sequence which encodes an *ANT1* polypeptide having at least 70% sequence identity to the amino acid sequence presented as SEQ ID NO:2.
2. The polynucleotide of Claim 1 comprising a nucleic acid sequence that hybridizes under high stringency conditions to the nucleic acid sequence presented as SEQ ID NO:1, or the complement or a fragment thereof.
3. The polynucleotide of Claim 1 wherein the *ANT1* polypeptide has at least 80% sequence identity to the amino acid sequence presented as SEQ ID NO:2.
4. The polynucleotide of Claim 1 wherein the *ANT1* polypeptide has at least 90% sequence identity to the amino acid sequence presented as SEQ ID NO:2.
5. The polynucleotide of Claim 1 wherein the *ANT1* polypeptide has the amino acid sequence presented as SEQ ID NO:2.
6. The polynucleotide of Claim 1 comprising the nucleic acid sequence presented as SEQ ID NO:1, or the complement thereof.
7. A plant transformation vector comprising an isolated polynucleotide of Claim 1.
8. A transgenic plant cell comprising the vector of Claim 7.
9. A method of producing an *ANT1* phenotype in a plant, said method comprising introducing into progenitor cells of the plant a plant transformation vector according to claim 7 and growing the transformed progenitor cells to produce a transgenic plant, wherein said polynucleotide sequence is expressed and said transgenic plant exhibits an *ANT1* phenotype.
10. A plant obtained by a method of Claim 9.
11. A plant part obtained from a plant according to Claim 10.

12. A method of selecting a transformed plant comprising a first polynucleotide comprising the steps of:
- (a) introducing into progenitor cells of the plant a plant transformation vector comprising the first polynucleotide and an *ANTI* polynucleotide according to Claim 1, and
  - (b) growing the progenitor cells to produce a plant that displays the *ANTI* phenotype, wherein the plant that displays the *ANTI* phenotype is selected as a transformed plant that also comprises the first polynucleotide.

## SEQUENCE LISTING

<110> Exelixis Plant Sciences, Inc.

<120> IDENTIFICATION AND CHARACTERIZATION OF AN ANTHOCYANIN MUTANT  
(ANT1) IN TOMATO

<130> EP01-002C-PC

<150> US 60/244,685

<151> 2000-10-30

<160> 5

<170> PatentIn version 3.1

<210> 1

<211> 825

<212> DNA

<213> Lycopersicon esculentum

<400> 1

atgaacagta catctatgtc ttcattggga gtgagaaaag gttcatggac tgatgaagaa	60
gattttcttc taagaaaatg tattgataag tatggtgaag gaaaatggca tcttggtccc	120
ataagagctg gtctgaatag atgtcggaaa agttgtagat tgaggtggct gaattatcta	180
aggccacata tcaagagagg tgactttgaa caagatgaag tggatctcat tttgaggctt	240
cataagctct taggcaacag atggtcactt attgctggta gacttcccg aaggacagct	300
aacgatgtga aaaactattg gaacactaat cttctaagga agttaatac tactaaaatt	360
gttcctcgcg aaaagattaa caataagtgt ggagaaatta gtactaagat tgaaattata	420
aaacctcaac gacgcaagta tttctcaagc acaatgaaga atgttacaaa caataatgta	480
attttggacg aggaggaaca ttgcaaggaa ataantaagt agaaacaaac tccagatgca	540
tcgatggaca acgtagatcc atgggtggata aatttactgg aaaattgcaa tgacgatatt	600
gaagaagatg aagaggttgt aattaattat gaaaaaacac taacaagttt gttacatgaa	660
gaaatatcac caccattaaa tattggtgaa ggtaactcca tgcaacaagg acaaataagt	720
catgaaaatt ggggtgaatt ttctcttaat ttaccacca tgcaacaagg agtacaaaat	780
gatgattttt ctgctgaaat tgacttatgg aatctacttg attaa	825

<210> 2

<211> 274

<212> PRT

<213> Lycopersicon esculentum

<400> 2

Met	Asn	Ser	Thr	Ser	Met	Ser	Ser	Leu	Gly	Val	Arg	Lys	Gly	Ser	Trp
1				5				10					15		

Thr Asp Glu Glu Asp Phe Leu Leu Arg Lys Cys Ile Asp Lys Tyr Gly  
20 25 30

Glu Gly Lys Trp His Leu Val Pro Ile Arg Ala Gly Leu Asn Arg Cys  
35 40 45

Arg Lys Ser Cys Arg Leu Arg Trp Leu Asn Tyr Leu Arg Pro His Ile  
50 55 60

Lys Arg Gly Asp Phe Glu Gln Asp Glu Val Asp Leu Ile Leu Arg Leu  
65 70 75 80

His Lys Leu Leu Gly Asn Arg Trp Ser Leu Ile Ala Gly Arg Leu Pro  
85 90 95

Gly Arg Thr Ala Asn Asp Val Lys Asn Tyr Trp Asn Thr Asn Leu Leu  
100 105 110

Arg Lys Leu Asn Thr Thr Lys Ile Val Pro Arg Glu Lys Ile Asn Asn  
115 120 125

Lys Cys Gly Glu Ile Ser Thr Lys Ile Glu Ile Ile Lys Pro Gln Arg  
130 135 140

Arg Lys Tyr Phe Ser Ser Thr Met Lys Asn Val Thr Asn Asn Asn Val  
145 150 155 160

Ile Leu Asp Glu Glu Glu His Cys Lys Glu Ile Ile Ser Glu Lys Gln  
165 170 175

Thr Pro Asp Ala Ser Met Asp Asn Val Asp Pro Trp Trp Ile Asn Leu  
180 185 190

Leu Glu Asn Cys Asn Asp Asp Ile Glu Glu Asp Glu Glu Val Val Ile  
195 200 205

Asn Tyr Glu Lys Thr Leu Thr Ser Leu Leu His Glu Glu Ile Ser Pro  
210 215 220

Pro Leu Asn Ile Gly Glu Gly Asn Ser Met Gln Gln Gly Gln Ile Ser  
225 230 235 240

His Glu Asn Trp Gly Glu Phe Ser Leu Asn Leu Pro Pro Met Gln Gln  
245 250 255

Gly Val Gln Asn Asp Asp Phe Ser Ala Glu Ile Asp Leu Trp Asn Leu  
260 265 270



Leu Asp

<210> 3  
 <211> 10078  
 <212> DNA  
 <213> pAG3202

<400> 3  
 tcccttttagt gagggttaat tccgagcttg gcgtaatcat ggcatagct gtttcctgtg 60  
 tgaaattgtt atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa 120  
 gcctgggggtg cctaagtgtg gagctaactc acattaattg cgttgcgctc actgcccgtc 180  
 ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga 240  
 ggcgggtttgc gtattgggag ctcttccgct tcctcgtcga ctgactcgtc gcgctcggtc 300  
 gttcggctgc ggcgagcggc atcagctcac tcaaaggcgg taatacgggt atccacagaa 360  
 tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt 420  
 aaaaaggccg cgttgctggc gtttttccat aggtccgcc cccctgacga gcatcacaaa 480  
 aatcgacgct caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcgttt 540  
 cccctggaa gctccctcgt gcgctctcct gttccgacce tgcgcttac cggataacct 600  
 tccgcctttc tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc 660  
 agttcgggtg aggtcgttcg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc 720  
 gaccgctgcg ccttatccgg taactatcgt cttgagtcca acccggttag acacgactta 780  
 tcgccactgg cagcagccac tggtaacagg attagcagag cgaggatatgt aggcgggtgt 840  
 acagagttct tgaagtgggt gcctaactac ggctacacta gaaggacagt atttggtatc 900  
 tgcgctctgc tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa 960  
 caaaccaccg ctggtagcgg tgggtttttt gtttgcaagc agcagattac gcgcagaaaa 1020  
 aaaggatctc aagaagatcc tttgatcttt tctacggggc ctgacgctca gtggaacgaa 1080  
 aactcacgtt aagggttttt ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt 1140  
 ttaaattaaa aatgaagttt taaatcaatc taaagtatat atgagtaaac ttggtctgac 1200  
 agttaccaat gcttaatcag tgaggcacct atctcagcga tctgtctatt tcgttcaccc 1260  
 atagttgcct gactccccgt cgtgtagata actacgatac gggagggcctt accatctggc 1320  
 cccagtgcgt caatgatacc gcgagacca cgtcacccgg ctccagattt atcagcaata 1380  
 aaccagccag ccggaagggc cgagcgcaga agtgggtcctg caactttatc cgcctccatc 1440  
 cagtctatta attgttgccg ggaagctaga gtaagtagtc cgcagtttaa tagtttgcgc 1500

aacgttggtg ccattgctac aggcacgtg gtgtcacgct cgtcgtttgg tatggcttca	1560
ttcagctccg gttcccaacg atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa	1620
gcggttagct ccttcgggtcc tccgatcgtt gtcagaagta agttggccgc agtggttatca	1680
ctcatgggta tggcagcact gcataattct cttactgtca tgccatccgt aagatgcttt	1740
tctgtgactg gtgagtactc aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt	1800
tgtctttgcc cggcggtcaat acgggataat accgcgccac atagcagaac tttaaaagtg	1860
ctcatcattg gaaaacgttc ttcggggcga aaactctcaa ggatcttacc gctgttgaga	1920
tccagttcga tgtaaccac tcgtgcaccc aactgatctt cagcatcttt tactttcacc	1980
agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg caaaaaaggg aataagggcg	2040
acacggaaat gttgaatact catactcttc ctttttcaat attattgaag catttatcag	2100
ggttattgtc tcatgagcgg atacatatct gaatgtattt agaaaaataa acaaataggg	2160
gttcgcgcga catttccccg aaaagtgcc cctgggaaat tgtaaacgtt aatattttgt	2220
taaaattcgc gttaaatttt tgtaaatca gtcattttt taaccaatag gccgaaatcg	2280
gcaaaatccc ttataaatca aaagaataga ccgagatagg gttgagtgtt gttccagttt	2340
ggaacaagag tccactatta aagaacgtgg actccaacgt caaagggcga aaaaccgtct	2400
atcagggcga tggcccacta cgtgaacat caccctaacc aagttttttg gggtcgaggt	2460
gccgtaaagc actaaatcgg aaccctaag ggagccccg atttagagct tgacggggaa	2520
agccggcgaa cgtggcgaga aaggaaggga agaaagcgaa aggagcgggc gctagggcgc	2580
tggcaagtgt agcggtcacg ctgcgcgtaa ccaccacacc cgccgcgctt aatgcgcgc	2640
tacagggcgc gtcgcgccat tcgccattca ggctgcgcaa ctggtgggaa gggcgatcgg	2700
tgcgggcctc ttcgctatta cgccagctgg cgaaaggggg atgtgctgca aggcgattaa	2760
gttgggtaac gccagggttt tcccagtcac gacgttgtaa aacgacggcc agtgaattgt	2820
aatacgactc actatagggc gaattggagc tccaccgcgg tggcgggcgc tctagaacta	2880
gtggatcccc aacatgggtg agcacgacac tctcgtctac tccaagaata tcaaagatac	2940
agtctcagaa gaccagaggg ctattgagac ttttcaacaa agggtaatat cgggaaacct	3000
cctcggtatc cattgccag ctatctgtca cttcatcgaa aggacagtag aaaaggaaga	3060
tggcttctac aaatgccatc attgcgataa aggaagggt atcgttcaag atgcctctac	3120
cgacagtggc cccaaagatg gacccccacc cagcaggaac atcgtggaaa aagaagacgt	3180
tccaaccacg tcttcaaagc aagtggattg atgtgatatc tagatcccca acatgggtgga	3240
gcacgacact ctcgtctact ccaagaatat caaagataca gtctcagaag accagagggc	3300
tattgagact tttcaacaaa gggtaatatc gggaaacctc ctcggtatcc attgcccage	3360

tatctgtcac	ttcatcgaaa	ggacagtaga	aaaggaagat	ggcttctaca	aatgccatca	3420
ttgcgataaa	ggaaaggcta	tcggtcaaga	tgcctctacc	gacagtgggc	ccaaagatgg	3480
acccccaccc	acgaggaaca	tcggtgaaaa	agaagacgtt	ccaaccacgt	cttcaaagca	3540
agtggttga	tgtgatctct	agatcccca	catggtggag	caegacactc	tctgtacttc	3600
caagaatctc	aaagatacag	tctcagaaga	ccagagggct	attgagactt	ttcaacaaag	3660
ggtaatatcg	ggaaacctcc	tgggattcca	ttgcccagct	atctgtcact	tcacgaaaag	3720
gacagtagaa	aaggaagatg	gcttctacaa	atgccatcat	tgcgataaag	gaaaggctat	3780
cgttcaagat	gcctctaccg	acagtgggtc	caaagatgga	ccccaccca	cgaggaacat	3840
cgtggaaaaa	gaagacgttc	caaccacgtc	ttcaaagcaa	gtggattgat	gtgatattcta	3900
gatccccaac	atggtggagc	acgacactct	cgtctactcc	aagaatatca	aagatacagt	3960
ctcagaagac	cagagggcta	ttgagacttt	tcaacaaagg	gtaatatcgg	gaaacctcct	4020
cggattccat	tgcccagcta	tctgtcactt	catcgaaagg	acagtagaaa	aggaagatgg	4080
cttctacaaa	tgccatcatt	gcgataaagg	aaaggtctatc	gttcaagatg	cctctaccga	4140
cagtgggtccc	aaagatggac	ccccaccac	gaggaacatc	gtggaaaaag	aagacgttcc	4200
aaccacgtct	tcaaagcaag	tggattgatg	tgatatctag	atccgaaact	atcagtgttt	4260
gacaggatat	attggcgggt	aaacctaaga	gaaaagagcg	tttattagaa	taatcggata	4320
tttaaaaggg	cgtgaaaagg	tttatccgtt	cgtccatttg	tatgtgcatg	ccaaccacag	4380
ggttccccctc	gggagtgttt	ggcattccgt	gcgataatga	cttctgttca	accacccaaa	4440
cgtcggaaaag	cctgacgacg	gagcagcatt	ccaaaaagat	cccttggttc	gtctgggtcg	4500
gctagaagggt	cgagtgggct	gctgtggctt	gatccctcaa	cgcggtcgcg	gacgtagcgc	4560
agcgccgaaa	aatcctcgat	cgcaaatccg	acgctgtcga	aaagcgtgat	ctgcttgctg	4620
ctcttttcggc	cgacgtcctg	gccagtcac	acgcgccaaa	gttccgtcac	aggatgatct	4680
ggcgcgagtt	gctggatctc	gccttcaatc	cgggtctgtg	gcgggaactc	cacgaaaata	4740
tccgaacgca	gcaagatatc	gcggtgcac	tcggtcttgc	ctgggcagtc	gccgccgacg	4800
ccgttgatgt	ggacgccggg	cccgatcata	ttgtcgctca	ggatcgtggc	gttgtgttg	4860
tcggccggtg	ctgtcgtaat	gatatcggca	ccttcgaccg	cctgttccgc	agaggtgcag	4920
gcctcgatct	gaaacccgaa	ccgctggaga	ttgcgggagc	agcgagcagt	agcctcgggg	4980
tcgatgtcgt	aaagtcgtat	ccgatcgacg	ccgatcagcg	ccttgaaggc	caaagcctgg	5040
aactcacttt	gggcaccgtt	gccgatcagc	gccatcgtgc	gcgaatcttt	acggggccaga	5100
tactttgccc	cgatcgcgga	ggtcgcgggc	gttcgcaagg	ccgtcaggat	tgtcatttcc	5160
gacagcagca	gcggatagcc	gctatcgaca	tcggagagca	cgccgaacgc	ggttaccggc	5220

ctcttcatcg ggaatgcgcg cgaccttcag catcgccggc atgtccccct ggcggacggg 5280  
 aagtatccag ctcgagatcc ggattttgta gccctggccg acggccagca ggtaggcga 5340  
 caggctcatg ccggccgccc ccgccttttc ctcaatcgct cttcgttcgt ctggaaggca 5400  
 gtacaccttg ataggtgggc tgccttcctt ggttggett gtttcatcag ccatccgctt 5460  
 gccctcatct gttacgcgcg cggtagccgg ccagcctcgc agagcaggat tcccgttgag 5520  
 caccgccagg tgcgaataag ggacagtga gaaggaacac ccgctcgcgg gtgggcctac 5580  
 ttcacctatc ctgcccggct gacgccgttg gatacaccaa ggaaagtcta cacgaaccct 5640  
 ttggcaaaat cctgtatata gtgcgaaaaa ggatggatat accgaaaaaa tcgctataat 5700  
 gaccccgaa gagggttatg cagcggaaaa gcgctgcttc cctgctgttt tgtggaatat 5760  
 ctaccgactg gaaacaggca aatgcaggaa attactgaac tgaggggaca ggcgagagac 5820  
 gatgccaaag agctacaccg acgagctggc cgagtgggtt gaatcccgcg cggccaagaa 5880  
 gcgcggcgt gatgaggctg cggttgcgtt cctggcggtg agggcggatg tcgaggcggc 5940  
 gttagcgtcc ggctatgcgc tcgtcaccat ttgggagcac atgcgggaaa cggggaagg 6000  
 caagttctcc tacgagacgt tccgctcgca cgccaggcgg cacatcatar gccaaagccc 6060  
 ccgatgtgcc cgcaccgcag gccaaaggctg cggaaccgc gccggcacc aagacgcgg 6120  
 agccacggcg gccgaagcag gggggcaagg ctgaaaagcc ggcccccgct gcggccccga 6180  
 ccggcttcac yttcaacca acaccggaca aaaaggatct agcgtggact caaggctctc 6240  
 gcgaatggct cgcgttgga actttcattg acacttgagg ggcaccgcag ggaaattcty 6300  
 gtcttgca gaaccggcta tgcgtgctg cgcatcgagc ctgcgccctt ggcttgcttc 6360  
 gccctctcc gcgtcgctac ggggcttcca gcgccttcc gacgtcacc gggctggttg 6420  
 ccctcgccgc tgggctggcg gccgttctat ggccttgcaa acgcgccaga aacgccgtcg 6480  
 aagccgtgtg cgagacaccg cggccgccgg cgttgkggat acctmgcgg aaacttgcc 6540  
 ctactgaca gatgaggggc ggacgttgac acttgagggg ccgactcacc cggcgcgcg 6600  
 ttgacagatg aggggcaggc tcgatttcgg ccggcgacgt ggagctggcc agcctcgaa 6660  
 atcggcga aaacgctgatt ttacgcgagt tccccacaga tgatgtggac aagcctgggg 6720  
 ataagtcccc tgcggtattg acacttgagg ggcgcgacta ctgacagatg aggggcgca 6780  
 tccttgacac ttgaggggca gagtgtgac agatgagggg cgcacctatt gacatttgag 6840  
 gggctgtcca caggcagaaa atccagcatt tgcaaggggt tccgccggtt ttccggccac 6900  
 cgctaacctg tcttttaacc tgcttttaaa ccaatattta taaaccttgt ttttaaccag 6960  
 ggtgcgccc tgtgcgctg accgcgcacg ccgaaggggg gtgcccccc ttctcgaacc 7020  
 ctcccgccc gctaacgcgg gcctcccatc ccccagggg ctgcgccct cggccgcgaa 7080

cggcctcacc ccaaaaatgg cagcgctggc agtccttgsc wwwgcykgsa tcggggcmst	7140
rmsggratgg gcgateagcc cgagcgcgac gcccgggaagc attgacgtgc cgcaggtgct	7200
ggcatcraca ttcagcgacc aggtgccggg cagtgagggc ggcggcctgg gtggcggcct	7260
gcccttcaact tcggccgtcg gggcattcac ggacttcacg gcggggccgg caatttttac	7320
cttgggcatt cttggcatak tggctcgggg tgccgtgctc gtgttcgggg gtgaattgca	7380
agctagcttg cttggctggt cccgtaccgt gaacgtcggc tcgattgtac ctgcgttcaa	7440
atactttgcg atcgtgttgc gcgcctgccc ggtgcgtcgg ctgatctcac ggatcgactg	7500
cttctctcgc aacgccatcc gacggatgat gtttaaaagt cccatgtgga tcactccgtt	7560
gccccgtcgc tcaccgtggt ggggggaagg tgcacatggc tcagttctca atggaaatta	7620
tctgcctaac cggctcagtt ctgcgtagaa accaacatgc aagctccacc gggtgcaaa	7680
cggcagcggc ggcaggatat attcaattgt aaatggcttc atgtccggga aatctacatg	7740
gatcagcaat gagtatgatg gtcaatatgg aggaaaagaa agagtaatta ccaatttttt	7800
ttcaattcaa aaatgtagat gtccgcagcg ttattataaa atgaaagtac attttgataa	7860
aacgacaaat tacgatccgt cgtatttata ggcgaagca ataaacaaat tattctaatt	7920
cggaaatctt tatttcgacg tgtctacatt caggtccaaa tgggggctta gatgagaaac	7980
ttcacgatcg atatctagat ctgcgactcg agatctagat atcgataagc ttgcatgcct	8040
gcagcggaat tcccatcttg aaagaaatat agtttaaata tttattgata aaataacaag	8100
tcaggatatta tagtccaagc aaaaacataa atttattgat gcaagtttaa attcagaaat	8160
atttcaataa ctgattatat cagctggtag attgccgtag atgaaagact gagtgcgata	8220
ttatgtgtaa tacataaatt gatgatatag ctagcttagc tcacggggg atccgtcgaa	8280
gctagcttgg gtcccgtcga gaagaactcg tcaagaaggc gatagaaggc gatgcgctgc	8340
gaatcgggag cggcgatacc gtaaagcacg aggaagcggg cagcccatte gccgccaagc	8400
tcttcagcaa tatcacgggt agccaacgct atgtcctgat agcgggtccgc cacacccagc	8460
cggccacagt cgatgaatcc agaaaagcgg ccattttcca ccatgatatt cggcaagcag	8520
gcacgcccat gggtcacgac gagatcctcg ccgtcgggca tgcgcgcctt gaggctggcg	8580
aacagttcgg ctggcgcgag cccctgatgc tcttcgtcca gatcatcctg atcgacaaga	8640
ccggcttcca tccgagtacg tgctcgtcgt atgcgatgtt tcgcttggtg gtcgaatggg	8700
caggtagccg gatcaagcgt atgcagccgc cgcattgcat cagccatgat ggatacttct	8760
tcggcaggag caagggtgaga tgacaggaga tccctccccg gcacttcgcc caatagcagc	8820
cagtcccttc ccgcttcagt gacaacgtcg agcacagctg cgcaaggaaac gcccgtcgtg	8880
gceagecacg atagccgcgc tgcctcgtcc tgcagttcat tcagggcacc ggacaggctg	8940

```

gtcttgacaa aaagaaccgg gcgcccctgc gctgacagcc ggaacacggc ggcatcagag 9000
cagccgattg tctgttgtyc ccagtcatac ccgaatagcc tctccacca agcggccgga 9060
gaacctgcgt gcaatccatc ttgttcaatc caagctccca tgggcccctt tcagtttgac 9120
tttggcgata gagggaggat cagttttata tgcagaaatg cgattgaatc taccgcgcaa 9180
agctactcag atacagacca actacttcga acgacatgtc gttgtaaagc taaataatct 9240
agttttgatt tcctgagcta atgttgatta gtagttagtg acttggtact gtaaacgcta 9300
acagtgaagt ggacgtgccg gagtgggtat cgggcatgac tagaaatcat gtgggctctg 9360
cgcgaggtt cgaatcctgc cgtccacgtc tttttcatgg gggtatcctt tttattgaga 9420
ttcgaaagca atttttggat ttggttggta atgcctcgag gccctcaaga aataagagtg 9480
aagaatctga caatatgtag tttagatatg taccctttga gtccataagt tttgaagttt 9540
gcatttccca tgggtatgta taagttgatg cacaattttg gttgcttttt attttagaaa 9600
aattacagga accgacaaaa ttgattataa acgtgagaaa tttatgtgtt taaatatgta 9660
gattgttttc tgactatata caatataaat tcagaatttg aatgtacttt aaaattaagt 9720
aagaaaaatt taggtagctc tcgcatatta tacctcaaag ttgggtataa attcgtaatt 9780
ttgtatactt caaaagtagt taatggaaac ataatagtat taatcagcat caaactcaac 9840
cataagttaa tgattacaac aaagcaattg tagcatcatt gtggctcgtc cggagtggtt 9900
atcgggcatg actagaaatc atgtgggctt tgcccgcgca ggttcgaatc ctgccgacca 9960
cgttttttta tgacatcatt tttaatcttt ttctcaaatg gaccaccatt aagcctaate 10020
ctgctgggtc tgggctaate atctcaatta agcttcccg gaaggtagca gcttttgt 10078

```

```

<210> 4
<211> 1012
<212> DNA
<213> Lycopersicon esculentum

```

```

<400> 4
atgaacagta catctatgtc ttcattggga gtgagaaaag gttcatggac tgatgaagaa 60
gattttcttc taagaaaatg tattgataag tatggtgaag gaaaatggca tcttgttccc 120
ataagagctg gtaactatta aattaactat caggttattt ttatttgtct ttctgtctca 180
ttttatttga cgttattacg aatatcatct gaaaatgtac gtgcaggtct gaatagatgt 240
cggaaaagtt gtagattgag gtggctgaat tatctaaggc cacatatcaa gagaggtgac 300
tttgaacaag atgaagtggg tctcattttt aggcttcata agctcttagg caacaggcat 360
gcaagtttat gttttgacaa aatttgatta gtatatatta tatatactg tgactatttc 420
atctaaatgt tacgttattt tacgtagatg gtaacttatt gctgtagac tttccggaag 480
gacagctaac gatgtgaaaa actattggaa cactaatctt ctaaggaagt taaatactac 540

```

taaaattggt	cctcgcgaaa	agattaacaa	taagtgtgga	gaaattagta	ctaagattga	600
aattataaaa	cctcaacgac	gcaagtattt	ctcaagcaca	atgaagaatg	ttacaaacaa	660
taatgtaatt	ttggacgagg	aggaacattg	caaggaaata	ataagtgaga	aacaaactcc	720
agatgcatcg	atggacaacg	tagatccatg	gtggataaat	ttactggaaa	attgcaatga	780
cgatattgaa	gaagatgaag	aggttgtaat	taattatgaa	aaaacactaa	caagtttggt	840
acatgaagaa	atatcaccac	cattaaatat	tgggtgaagg	aactccatgc	aacaaggaca	900
aataagtcac	gaaaattggg	gtgaattttc	tcttaattta	ccacccatgc	aacaaggagt	960
acaaaaatgat	gatttttctg	ctgaaattga	cttatggaat	ctacttgatt	aa	1012

<210> 5  
 <211> 12241  
 <212> DNA  
 <213> pAG2370

<400> 5	
tgagcgtcgc	aaaggcgctc ggtcttgcct tgctcgtcgg tgatgtactt caccagctcc 60
gcgaagtcgc	tcttcttgat ggagcgcatg gggacgtgct tggcaatcac gcgcaccccc 120
cggccgtttt	agcggctaaa aaagtcattg ctctgcctc gggcggacca cgcccatcat 180
gaccttgcca	agctcgctct gcttctcttc gatcttcgcc agcagggcga ggatcggtggc 240
atcacccaac	cgcgccgtgc gcgggtcgtc ggtgagccag agtttcagca ggccgcccag 300
gcggcccagg	tcgccattga tgcgggccag ctgcgggacg tgctcatagt ccacgacgcc 360
cgtagttttg	tagccctggc gcagggccag caggtaggcc gacaggtca tgccggccgc 420
cgccgccttt	tcctcaatcg ctcttcgttc gtctggaagg cagtacacct tgataggtgg 480
gctgcccttc	ctggttggct tggtttcac agccatccgc ttgccctcat ctgttacgcc 540
ggcggtagcc	ggccagcctc gcagagcagg attcccgttg agcacgccca ggtgcgaata 600
agggacagtg	aagaaggaaac acccgctcgc ggggtgggct acttcaccta tcctgcccgg 660
ctgacgccgt	tggatacacc aaggaaagtc tacacgaacc ctttggcaaa atcctgtata 720
tcgtgcgaaa	aaggatggat ataccgaaaa aatcgctata atgacccga agcagggtta 780
tgcagcggaa	aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg 840
gcagggtcgg	aacaggagag cgcacgaggg agcttcagg gggaaacgcc tggatatctt 900
atagtcctgt	cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag 960
gggggcccgg	cctatggaaa aacgccagca acgcccctt ttacgggttc ctggcctttt 1020
gctggccttt	tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta 1080
ttaccgcctt	tgagtgaagt gataccgctc gccgcagccg aacgaccgag cgcagcgagt 1140

cagtgagcga ggaagcggaa gagcgccaga aggccgccag agaggccgag cgcggccgtg 1200  
 aggcttgga gctagggcag ggcataaaaa agcccgtagc gggctgctac gggcgtctga 1260  
 cgcggtggaa agggggaggg gatgttgtct acatggctct gctgtagtga gtgggttgcg 1320  
 etccggcagc ggtcctgac aatcgtcacc ctttctcggc ccttcaacgt tcctgacaa 1380  
 gagcctcctt ttccccaatc catcgacaat caccgcgagt ccctgctcga acgctgcgtc 1440  
 cggaccggct tcgtcgaagg cgtctatcgc gggccgcaac agcggcgaga ggggagcctg 1500  
 ttcaacgggtg ccgccgcgct cgcgggcac cgtgtcgcg gcctgctcct caagcacggc 1560  
 cccaacagtg aagtagctga ttgtcatcag cgcattgacg gcgtcccg cggaaaaacc 1620  
 cgcctcgcag aggaagcga gctgcgcgtc ggcggtttcc atctcgggtg cgcgggtcg 1680  
 cgtgccggca tggatgcgcg cgcctcgcg gtaggcgagc agcgcctgcc tgaagctgcg 1740  
 ggcattcccg atcagaaatg agcgcagtc gtcgtcggct ctccggcaccg aatgcgtatg 1800  
 attctccgcc agcatggctt cggccagtgc gtcgagcagc gcccgcttgt tcctgaagtg 1860  
 ccagtaaaagc gccggctgct gaaccccaaa ccgttcgcc agtttgctg tcgtcagacc 1920  
 gtctacgcg acctcgttca acaggtccag gggggcacgg atcactgtat tcggctgcaa 1980  
 ctttgtcatg cttgacactt tatcactgat aaacataata tgtccaccaa cttatcagtg 2040  
 ataaagaatc cgcgcgttca atcggaccag cggaggctgg tccggaggcc agacgtgaaa 2100  
 cccaacatac ccctgatcgt aattctgagc actgtcgcg tcgacgctgt cggcatcggc 2160  
 ctgattatgc cgggtctgcc gggcctcctg cgcgatctgg ttactcga cgcagtcacc 2220  
 gccactatg gcattctgct ggcgctgtat gcgttggtgc aatttgctg cgcacctgtg 2280  
 ctgggcgcgc tgteggatcg tttcggggcg cggccaatct tgctcgtctc gctggcggc 2340  
 gccagatctg gggaaacctg tggttggcat gcacatacaa atggacgaac ggataaacct 2400  
 tttcacgccc ttttaaatat ccgattattc taataaacgc tcttttctct taggtttacc 2460  
 cgccaatata tcctgtcaaa cactgatagt ttaactgaa ggcgggaaac gacaatctga 2520  
 tcatgagcgg agaattaagg gagtcacgtt atgaccccg ccgatgacgc gggacaagcc 2580  
 gttttacgtt tggaaactgac agaaccgcaa cggtgaagga gccactcagc cgatctgaat 2640  
 tcccgatcta gtaacataga tgacaccgcg cgcgataatt tatcctagtt tgcgcgctat 2700  
 attttgttt ctatcgcgta ttaaatgtat aattgcggga ctctaactcat aaaaacccat 2760  
 ctcataaata acgtcatgca ttacatgta attattacat gcttaacgta attcaacaga 2820  
 aattatatga taatcatcgc aagaccggca acaggattca atcttaagaa actttattgc 2880  
 caaatgtttg aacgatcggg gaaattcgcg agctcggtag ccgctctaga actagtggat 2940  
 ccccgggct gcaggaattc aaacttaca atttctctga acttgatcc tcagtacttc 3000



aaagaaaata gcttacacca aattttttct tgttttcaca aatgccgaac ttggttcctt	3060
atatagggaaa actcaagggc aaaaatgaca cggaaaaata taaaaggata agtagtgggg	3120
gataagattc ctttgtgata aggttacttt ccgcccttac attttccacc ttacatgtgt	3180
cctctatgtc tctttcacaa tcaccgacct tatcttcttc ttttcattgt tgcgtcagt	3240
gcttacgtct tcaagattct tttcttcgcc tgggtcttct ttttcaattt ctacgtattc	3300
ttcttcgtat tctggcagta taggatcttg tatctgtaca ttcttcattt ttgaacatag	3360
gttgcataatg tgccgcataat tgatctgctt cttgctgagc tcacataata cttccatagt	3420
ttttcccgta aacattggat tcttgatgct acatcttggg taattacctt ctcgtacca	3480
gcttaattga gatgattagc ccagaccag caggattagg cttaatgggtg gtccatttga	3540
gaaaaagatt aaaaatgatg tcataaaaaa acgtggtcgg caggattcga acctgcgcgg	3600
gcaaagccca catgatttct agtcatgccc gataaccact ccggcacgac cacaatgatg	3660
ctacaattgc tttgttgtaa tcattaactt atggttgagt ttgatgctga ttaatactat	3720
tatgtttcca ttaactactt ttgaagtata caaaattacg aatttataac caaatttgag	3780
gtataatatg cgagagctac ctaaaattttt ctacttaat tttaaagtae attcaaattc	3840
tgaatttata ttgtgtatag tcagaaaaca atctacatat ttaaacacat aaatttctca	3900
cgtttataat caattttgtc ggttcctgta atttttctaa aataaaaagc aaccaaattt	3960
gtgcatcaac ttattacata ccatgggaaa tgcaaacttc aaaacttatg gactcaaagg	4020
gtacatatct aaactacata ttgtcagatt cttcactctt atttcttgag ggcctcgagg	4080
cattaccaac caaatccaaa aattgcttct gaatctcaat aaaaaggata accccatgaa	4140
aaagacgtgg acggcaggat tcgaacctgc gcgcagagcc cacatgattt ctagtcatgc	4200
ccgataacca ctccggcacg tccacttcac tgtaacgtt tacagtaaca agtcaactaac	4260
tactaatcaa cattagctca ggaaatcaaa actagattat ttacatttac aacgacatgt	4320
cgttcgaagt agttggctctg tatctgagta gctttggcgg gtagattcaa tcgcatttct	4380
gcatataaaa ctgatcctcc ctctatcgcc aaagtcaaac tgaaaagggc cgggggcaag	4440
atatgggagc ttggattgaa caagatggat tgacgcagc ttctccggcc gcttgggtgg	4500
agaggctatt cggctatgac tgggcacaa agacaatcgg ctgctctgat gccgccgtgt	4560
tccggctgtc agcgcagggg gcgccggttc tttttgtcaa gaccgacctg tccgggtgcc	4620
tgaatgaact gcaggacgag gcagcgcggc tategtggct ggccaacgacg ggcgttcctt	4680
gcgcagctgt gctcgacgtt gtcactgaag cgggaaggga ctggctgcta ttgggcgaag	4740
tgccggggca ggatctctctg tcactctacc ttgctcctgc cgagaaagta tccatcatgg	4800
ctgatgcaat gcggcggtctg catacgcttg atccggctac ctgcccattc gaccaccaag	4860

cgaaacatcg catcgagcga gcacgtactc ggatggaagc cgytcttgtc gatcaggatg 4920  
 atctggacga agagcatcag gggctcgcgc cagccgaact gtctgccagg ctcaaggcgc 4980  
 gcatgcccga cggcgaggat ctctcgtga cccatggcga tgcctgcttg ccgaatatca 5040  
 tgggtgaaaa tggccgcttt tctggattca tcgactgtgg ccggttgggt gtggcggacc 5100  
 gctatcagga catagcgttg gctaccctg atattgctga agagcttggc ggcaaatggg 5160  
 ctgaccgctt cctcgtgctt tacggtatcg ccgctccga ttgcagcgc atcgcttct 5220  
 atcgcttct tgacgagttc ttctgacgat gagctaagct agctatatca tcaatttatg 5280  
 tattacacat aatatcgac tcagtcttct atctacggca atgtaccagc tgatataatc 5340  
 agttattgaa atatttctga atttaaactt gcatcaataa atttatgttt ttgcttggac 5400  
 tataatacct gacttggtat tttatcaata aatattttaa ctatatttct ttcaagatgg 5460  
 gaattaattc actggcgcgc gttttacaac gtcgtgactg ggaaaaccct ggcgttacc 5520  
 aacttaatcg ccttgacga catccccct tcgccagctg gcgtaatagc gaagaggccc 5580  
 gcaccgatcg ccttcccaa cagttgcgca gcctgaatgg cggcgctcc ttctgcttct 5640  
 ttcccttct ttctgccac gtctgccgc tttcccgctc aagctctaaa tcgggggctc 5700  
 cctttagggt tccgatttag tgctttacgg cactcgacc ccaaaaaact tgatttgggt 5760  
 gatggttcac gtagtgggc atcgccctga tagacggttt ttgccttct gacgttggag 5820  
 tccacgttct ttaatagtgg actctgttc caaactggaa caacactcaa ccctatctcg 5880  
 ggctattctt ttgatttata agggattttg ccgatttcgg aaccaccatc aaacaggatt 5940  
 ttgcctgct ggggcaaacc agcgtggacc gcttgcgca actctctcag ggccaggcgg 6000  
 tgaagggcaa tcagctgttg cccgtctcac tgggtgaaaag aaaaaccacc ccagtacatt 6060  
 aaaaacgtcc gcaatgtgtt attaagttgt ctaagcgtca atttgtttac accacaatat 6120  
 atcctgccac cagccagcca acagctcccc gaccggcagc tcggcacaaa atcaccactc 6180  
 gatacaggca gcccatcagt ccgggacggc gtcagcggga gagccgttgt aaggcggcag 6240  
 actttgctca tgttaccgat gctattcgga agaacggcaa ctaagctgcc gggtttgaaa 6300  
 cacggatgat ctgcggagg gtagcatgtt gattgtaacg atgacagagc gttgctgcct 6360  
 gtgatcaaat atcatctccc tcgcagagat ccgaattatc agccttctta ttcatttctc 6420  
 gcttaaccgt gacaggctgt cgatcttgag aactatgccg acataatagg aaatcgctgg 6480  
 ataaagccgc tgaggaagct gagtggcgt atttctttag aagtgaacgt tgacgatatc 6540  
 aactccccta tccattgctc accgaatggt acaggtcggg gaccggaagt tccgactgtc 6600  
 ggcctgatgc atccccggt gatcgacccc agatctgggg ctgagaaagc ccagtaagga 6660  
 aacaactgta ggttcgagtc gcgagatccc ccggaaccaa aggaagtagg ttaaaccgcg 6720

tccgatcagg cegagccacg ccaggccgag aacattgggt cctgtaggca tcgggattgg	6780
cggatcaaac actaaagcta ctggaacgag cagaagtcct cgggcccga gttgccaggc	6840
ggtaaagggt agcagaggca cgggagggtt ccacttgccg gtcagcacgg ttccgaacgc	6900
catggaaacc gcccccgcca ggcccgtgc gacgccgaca ggatctagcg ctgcgtttgg	6960
tgtcaacacc aacagcgcca cggccgagc tccgcaaata gccccagga ccgccatcaa	7020
tcgtatcggg ctacctagca gagcggcaga gatgaacacg accatcagcg gctgcacagc	7080
gcctaccgtc gccgcgaccc cggccggcag gcggtagacc gaaataaaca acaagctcca	7140
gaatagcgaa atattaagtg gcgcgaggat gaagatgcgc atccaccaga ttcccgttgg	7200
aatctgtcgg acgatcatca cgagcaataa acccgccggc aacgcccga gcagcatacc	7260
ggcgacccct cggcctcgct gttcgggctc cagcaaacg cgggacagat gcgccttgtg	7320
agcgctcctt gggcgctcct cctgtttgaa gaccgacagc ccaatgatct cggcgtcgat	7380
gtaggcgccg aatgccacgg catctcgcaa ccgttcagcg aacgcctcca tgggcttttt	7440
ctcctcgtgc tcgtaaacgg acccgaacat ctctggagct ttcttcaggg ccgacaatcg	7500
gatctcggcg aaatcctgca cgtcggccgc tccaagccgt cgaatctgag ccttaatcac	7560
aattgtcaat ttaatacctc tgtttatcgg cagttcgtag agcgcgccgt gcgtcccag	7620
cgatactgag cgaagcaagt gcgtcgagca gtgcccgtt gttcctgaaa tgccagtaaa	7680
gcgctggctg ctgaaccccc agccggaact gacccacaaa ggccctagcg tttgcaatgc	7740
accaggteat cattgaccca ggcgtgttcc accaggccgc tgcctcgcaa ctcttcgcag	7800
gcttcgccga cctgctcgcg ccacttcttc acgcgggtgg aatccgatcc gcacatgagg	7860
cgggaagggtt ccagcttgag cgggtacggc tcccgggtgc agctgaaata gtcgaacatc	7920
cgtcggggcg tcggcgacag cttgcggtac ttctccata tgaatttcgt gtagtggtcg	7980
ccagcaaaaca gcacgacgat ttcctcgtcg atcaggacct ggcaacggga cgttttcttg	8040
ccacggtcca ggacgcggaa gcggtgcagc agcgacaccg attccagggt cccaacgcgg	8100
tcggacgtga agcccatcgc cgtcgctgt aggcgcgaca ggcattcctc ggccttcgtg	8160
taataccggc cattgatcga ccagcccagg tcctggcaaa gctcgtagaa cgtgaagggt	8220
atcggctcgc cgataggggt gcgcttcgcg tactccaaca cctgctgcca caccagttcg	8280
tcacgtcggg cccgcagctc gacgccggtg taggtgatct tcacgtcctt gttgacgtgg	8340
aaaatgacct tgttttgag cgctcgcgc gggattttct tgttgccggt ggtgaacagg	8400
gcagagcggg ccgtgtcgtt tggcatcgct cgcacgtgt ccggccacgg cgcaatatcg	8460
aacaaggaaa gctgcatttc cttgatctgc tgcttcgtgt gtttcagcaa cgggcctgc	8520
ttggcctcgc tgacctgttt tgccaggctc tcgcccggcg ttttctgctt cttggctcgc	8580

ctcaacacga ttttacgtca cttaaaaaac tcagggccga gtcggtaacc tcgcgcatac 10500  
 agccgggagc tgacgtcatc gtctgcgcgg aaatggacga acagtggggc tatgtcgggg 10560  
 ctaaatacgc ccagcgctgg ctgttttacg cgtatgacag tctccggaag acggttggtg 10620  
 cgcacgtatt cggatgaacgc actatggcga cgctggggcg tcttatgagc ctgctgtcac 10680  
 cctttgacgt ggtgatatgg atgacggatg gctggccgct gtatgaatcc cgcctgaagg 10740  
 gaaagctgca cgtaatcagc aagcgatata cgcagcgaat tgagcggcat aacctgaatc 10800  
 tgaggcagca cctggcacgg ctgggacgga agtcgctgtc gttctcaaaa tcgggtggagc 10860  
 tgcattgaca agtcattcgg cattatctga acataaaaca ctatcaataa gttggagtca 10920  
 ttacccaatt atgatagaat ttacaagcta taaggttatt gtctctgggt tcaagcatta 10980  
 gtccatgcaa gtttttatgc tttgcccatt ctatagatat attgataagc gcgctgccta 11040  
 tgccttgccc cctgaaatcc ttacatacgg cgatatcttc tatataaaag atatattatc 11100  
 ttatcagtat tgtcaatata ttcaaggcaa tctgcctcct catcctcttc atcctcttcg 11160  
 tcttggtagc tttttaaata tggcgcttca tagagtaatt ctgtaaaggc ccaattctcg 11220  
 ttttcatacc tcgggtataat cttacctatc acctcaaagc gtctcgctggg tttatcgac 11280  
 ccccgaacac gagcacggca cccgcgacca ctatgccaaag aatgcccaag gtaaaaattg 11340  
 ccggcccccgc catgaagtcc gtgaatgcc cgacggccga agtgaagggc aggccgccac 11400  
 ccaggccgcc gccctcactg cccggcacct ggctcgctgaa tgcgatgcc agcacctgcg 11460  
 gcacgtcaat gcttcggggc gtcgcgctcg ggctgatcgc ccatcccggt actgccccga 11520  
 tcccggcaat ggcaaggact gccagcgctg ccatttttgg ggtgaggccg ttcgcggccg 11580  
 aggggcgcag cccctggggg gatgggaggc ccgcgttagc gggccgggag ggttcgagaa 11640  
 gggggggcac cccctctcgg cgtgcgcggt cacgcgcaca gggcgagcc ctggttaaaa 11700  
 acaaggttta taaatattgg tttaaaagca ggttaaaaga cagggttagc gtggccgaaa 11760  
 aacgggcgga aacccttgca aatgctggat tttctgcctg tggacagccc ctcaaagtgc 11820  
 aataggtgcg cccctcatct gtcagcactc tgccctcaa gtgtcaagga tcgcgccct 11880  
 catctgtcag tagtcgcgcc cctcaagtgt caataccgca gggcacttat ccccaggctt 11940  
 gtccacatca tctgtgggaa actcgcgtaa aatcaggcgt tttcgccgat ttgcgagget 12000  
 ggccagctcc acgtcgccgg ccgaaatcga gcctgcccc catctgtcaa cgcgcgccc 12060  
 ggtgagtcgg cccctcaagt gtcaacgtcc gccctcatc tgcagtgag ggccaagttt 12120  
 tccgcgaggt atccacaacg ccggcggccg cgggtgtctc cacacggctt cgacggcggt 12180  
 tctggcgcgt ttgcagggcc atagacggcc gccagcccag cggcgagggc aaccagccc 12240  
 g 12241